GEOMETRY AND PROBABILITY
Math 545 - Fall 2001

Renato Feres

Contents

CHAPTER 1

An Informal Introduction to Brownian Motion

1. Stochastic Differential Equations

A "deterministic" system of ordinary differential equations is specified by a vector field. Say that X denotes a differentiable vector field on $\mathbb{R}^n$, possibly time dependent. Then a solution to the initial value problem

$$\dot{x} = X(t,x), x(0) = x_0 \tag{1.1}$$

is a function $x(t)$ that describes a differentiable curve passing through the point $x_0 \in \mathbb{R}^n$ whose derivative vector (velocity) is equal to $X(t, x(t))$ at each point $x(t)$ along the curve.

For example, consider the initial value problem

$$\begin{aligned} \ddot{z} + a\dot{z} + bz &= f(t) \\ z(0) &= z_1 \\ \dot{z}(0) &= z_2. \end{aligned}$$

The solution might represent, for example, the motion of a mass attached to a spring, with position and velocity at time 0 specified by the two initial conditions. If one defines

$$x = \begin{pmatrix} z \\ \dot{z} \end{pmatrix},\ A = \begin{pmatrix} 0 & 1 \\ -b & -a \end{pmatrix},\ F(t) = \begin{pmatrix} 0 \\ f(t) \end{pmatrix},\ X(t,x) = Ax + F(t),\ x_0 = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$$

then the second order equation can be expressed by the first order system and initial condition 1.1. The solution can be written as

$$x(t) = e^{tA}x_0 + e^{At}\int_0^t e^{-As}F(s)ds. \tag{1.2}$$

In a more realistic description of the motion, $x(0)$ might be regarded as a random quantity, taking a range of values according to some given probability distribution. Furthermore, the forces acting on the mass may also have a random component. For example, if we imagine that the mass is very small and that the motion can be affected to some small degree by random motion of air molecules, then it would be appropriate to describe the total force acting on the mass as also being a random - or "noisy" - quantity. We could then try to represent the problem by an equation of the type

$$\dot{x} = X(t,x) + N(t) \tag{1.3}$$

where $N(t)$ corresponds to the noise, ("random component") of the driving vector field. If N were an ordinary continuous function, the solution to 1.3 would be

$$x(t) = x_1(t) + e^{At}\int_0^t e^{-As}N(s)ds, \tag{1.4}$$

where $x_1(t)$ is the solution to 1.1.

We are faced with the issue of defining $N(t)$. Although a random process $N(t)$ having the properties one would like to assign to "noise" cannot be easily defined, it turns out that $B(t) = \int_0^t N(s)ds$ can be given a relatively simple mathematical interpretation and serve as the foundation for a calculus with noise.

Our first order of business will be to obtain a mathematical characterization of $B(t)$ (first on $\mathbb{R}^n$ and, later, on general Riemannian manifolds).

Some of the desirable properties that $B(t)$ should have are:

(1) Unpredictability: the increments $\Delta B(t) = B(t+\Delta t) - B(t)$ are completely unpredictable, that is, $\Delta B(t)$ is independent of its past $\{B(s) : s \leq t\}$;
(2) Stationarity: $B(t)$ is stationary, in the sense that the probability distribution of $\Delta B(t)$ does not depend on t;
(3) Continuity: if we imagine that the erratic, unpredictable behaviour of $B(t)$ is the result of a large number of relatively weak independent factors, then it is reasonable to expect that $B(t)$ does not have finite jumps, that is, that $B(t)$ is continuous in some sense.

It will be seen that the previous properties characterize the so called *Brownian motion*, or *Wiener*, process.

It is natural to ask whether one can find a probability density $\rho(t,x)$ that gives the probability of finding $x(t)$ in a region $K \subset \mathbb{R}^2$ as an integral

$$P(x(t) \in K) = \int_K \rho(t,x)dx.$$

It will be seen later that ρ is a solution of a parabolic (deterministic) partial differential equation - a diffusion equation.

2. The Microscopic View

We would like to take here a closer look at noise from a microscopic viewpoint, illustrating the main points with a simple mechanical example.

Suppose that a body of mass M, whose motion we would like to describe, has the shape of a parallelepiped and moves freely, without friction, on a horizontal rail. The body will be imagined to have sufficiently small mass for its motion to be affected by the thermal motion of gas molecules around it.

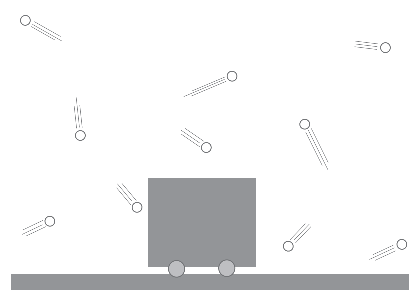
To study the motion of the body it will be convenient to keep in mind a number of different time scales: First, there is the "macroscopic scale," in which some observable displacement can be detected (say, a few minutes.) Next, we may consider the "calculus scale". Quantities that we will write as Δt (such as when we approximate the derivative of a differentiable function by the quotient $(f(t+\Delta t) - f(t))/\Delta t)$ will belong to this scale. Our Δt might be of the order of, say, hundredths of a second. Finally, we will distinguish the "microscopic scale" measuring the typical time interval between collisions of gas molecules with the body. The mean time between collisions will be denoted by τ, and will depend on the gas density and mean velocity.

In order to be a little more precise it will be convenient to define

$$\gamma := \frac{1 - \frac{m}{M}}{1 + \frac{m}{M}},$$

where m is the mass of individual gas molecules. We assume that we can choose times scales so that the following hold:

(1) Δt should be sufficiently bigger than τ that $\Delta t/\tau$ will come "very close" to being the number of collisions during an interval $[t, t + \Delta t]$. (The law of large numbers should lie behind this property.)
(2) The ratio m/M is so small that γ^i is "very close" to 1 for all $i \leq [\Delta t/\tau]$.



The component of the molecules' velocity along the direction of the rail will be written v. It will be regarded as a random variable with mean 0 and mean square

$$\mathbb{E}(v^2) = \frac{kT}{m},$$

where k is a physical quantity known as the *Boltzmann constant* and T is the *temperature* of the gas. (The temperature T may be defined, up to constant, as the mean kinetic energy of gas molecules. This will be further discussed later on.)

The effect that a collision with a gas molecule will have on the body can be calculated as follows. Denoting by V, V', and v, v' the velocities (along the rail) of the body (capitalized) and molecule (lower case) before and after the collision, then we have

$$MV + mv = MV' + mv'$$
$$\frac{1}{2}MV^2 + \frac{1}{2}mv^2 = \frac{1}{2}MV'^2 + \frac{1}{2}mv'^2,$$

where the first equation describes conservation of momentum and the second describes concervation of energy. (we are assuming that the collisions are perfectly elastic.)

These equations are easly solved, giving:

$$v' = (1 + \gamma)V - \gamma v$$
$$V' = \gamma V + (1 - \gamma)v.$$

Now suppose that the initial velocity of the body is $V_0 = V(0)$, and that the collision times after $t = 0$ are $T_1 < T_2 < T_2 < \dots$. The horizontal components of the velocities of the colliding molecules at the respective times are $v_1, v_2, v_2, \dots$. We assume that v_i are independent identically distributed random variables (with mean 0 and mean square kT/m) and that $\tau_i = T_{i+1} - T_i$ are independent identically distributed random variables with mean τ.

Then, for $t \in [T_i, T_{i+1})$, and writing $V_i := V(T_i)$,

$$V_i = V(t) = \gamma V_{i-1} + (1-\gamma) v_i.$$

Iterating the previous equation we obtain, after n collisions,

$$V_n = \gamma^n V_0 + (1-\gamma) \sum_{i=0}^{n-1} \gamma^i v_{n-i}.$$

By the law of large numbers, the number of collisions up to time t is, approximately, $n = [t/\tau]$. Therefore, we write

$$V(t) = e^{-\alpha t} V_0 + (1-\gamma) \sum_{i=0}^{[\frac{t}{\tau}]-1} \gamma^i v_{[t/\tau]-i},$$

where $\alpha := (\ln \gamma^{-1})/\tau$.

The exponentially decaying term $e^{-\alpha t} V_0$ shows that after a little while any memory of the initial velocity is lost. From now on, we ignore that term by assuming that $V_0 = 0$.

We make now a key observation. Write $\Delta V_t := V(t + \Delta t) - V(t)$. Then

$$\begin{aligned}
\Delta V_t &= (1-\gamma) \left(\sum_{i=0}^{[\frac{t+\Delta t}{\tau}]-1} \gamma^i v_{[\frac{t+\Delta t}{\tau}]-i} - \sum_{i=0}^{[\frac{t}{\tau}]-1} \gamma^i v_{[\frac{t}{\tau}]-i} \right) \\
&= (1-\gamma) \left(\sum_{i=1}^{[\frac{t+\Delta t}{\tau}]} \gamma^{[\frac{t+\Delta t}{\tau}]-i} v_i - \sum_{i=1}^{[\frac{t}{\tau}]} \gamma^{[\frac{t}{\tau}]-i} v_i \right) \\
&= (1-\gamma) \left(\sum_{i=[\frac{t}{\tau}]+1}^{[\frac{t+\Delta t}{\tau}]} \gamma^{[\frac{t+\Delta t}{\tau}]-i} v_i - \sum_{i=1}^{[\frac{t}{\tau}]} \left(\gamma^{[\frac{t+\Delta t}{\tau}]-i} - \gamma^{[\frac{t}{\tau}]-i} \right) v_i \right).
\end{aligned}$$

We are assuming that the v_i are independent and identically distributed, with mean 0 and mean square $\nu^2 := kT/m$. In particular, $\mathbb{E}[v_i v_j] = 0$ whenever $i \neq j$. With that in mind, we write:

$$\begin{aligned}
\mathbb{E}[(\Delta V_t)^2] &= \nu^2 (1-\gamma)^2 \left(\sum_{i=[\frac{t}{\tau}]+1}^{[\frac{t+\Delta t}{\tau}]} \gamma^{2([\frac{t+\Delta t}{\tau}]-i)} - \sum_{i=1}^{[\frac{t}{\tau}]} \left(\gamma^{2([\frac{t+\Delta t}{\tau}]-i)} - \gamma^{2([\frac{t}{\tau}]-i)} \right) \right) \\
&= \nu^2 (1-\gamma)^2 \left(\sum_{i=0}^{[\frac{t+\Delta t}{\tau}]-[\frac{t}{\tau}]-1} \gamma^{2i} - \left(\gamma^{2([\frac{t+\Delta t}{\tau}]-[\frac{t}{\tau}])} - 1 \right) \sum_{i=0}^{[\frac{t}{\tau}]-1} \gamma^{2i} \right)
\end{aligned}$$

We now use the scale assumptions. A more precise statement of the assumptions that is actually used is the following. The quantities γ and $\Delta t/\tau$ are such that $(1-\gamma^i)^2$ is much smaller than $1-\gamma^2$ for all $i = 1, \ldots, [\Delta t/\tau] + 1$. (This is clearly possible since the limit of $(1-\gamma^i)^2/(1-\gamma^2)$ as $\gamma \to 1$ is 0.)

Under these approximations, the second sum is close to 0, γ^{2i} is close to 1 for all $i = 1, \ldots, [\Delta t/\tau] + 1$, and we have (disregarding small errors)

$$\mathbb{E}[(\Delta V_t)^2] = \nu^2 (1-\gamma)^2 \left([\frac{t+\Delta t}{\tau}] - [\frac{t}{\tau}] \right) = \text{Constant} \Delta t.$$

The above calculation actually shows that ΔV_t is close to

$$(1-\gamma)(\sum_{i=[\frac{t}{\tau}]+1}^{[\frac{t+\Delta t}{\tau}]} v_i),$$

a sum of independent identically distributed random variables with mean 0 and finite variance. Since, by our scale assumption, $\Delta t/\tau$ is big, the central limit theorem implies that ΔV_t is a centered Gaussian random variable, with variance $C\Delta t$.

Therefore, the random process $V(t)$ has the following properties: ΔV_t is a Gaussian random variable with mean 0, variance $C\Delta t$ and it is independent of $\{V(s) : s \leq t\}$. Later on, we will use precisely these properties to define Brownian motion.

Notice that in this physical situation it is the velocity process that is a Brownian motion, whereas the position process is obtained by integrating $V(t)$.

3. Temperature and the Maxwell-Boltzmann distribution

Consider a collection of d noninteracting point masses of equal mass m, moving freely inside a rectangular box B with solid walls. As each particle reaches a wall of the box, it bounces off according to the usual equal angles law. The total energy of the gas as a function of the velocities is

$$E(v_1, \dots, v_d) := \sum_{i=1}^{d} \frac{1}{2} m |v_i|^2.$$

It is assumed constant and equal to E. The mean energy per molecule will be written as

$$\frac{E}{d} =: \frac{3}{2} kT$$

where k is a constant independent of E and d (the Boltzmann constant) and T is the *temperature* of the gas. The phase space of the system (whose points represent the positions and velocities of the particles) is

$$(B \times \mathbb{R}^3)^d = B^d \times \mathbb{R}^{3d}.$$

Assuming that the total energy is constant and equal to E (the box is thermally insulated so that no energy exchange takes place with the outside), the part of phase space the gas may occupy is $\Omega := B^d \times S$, where S is the sphere in $\mathbb{R}^{3d}$ with center 0 and radius $R = \sqrt{3kT/m}\ d^{1/2} = cd^{1/2}$. We now make the assumption that the particles are distributed in Ω according to the uniform distribution. This means that P will be taken to be the normalized volume measure on Ω.

We would like to determine the probability that a velocity component of a given particle will fall in the interval $[a, b]$. Clearly, the position of the particle is immaterial, so the problem has the following geometric formulation. Let S denote the sphere in $\mathbb{R}^{3d}$ with center 0 and radius $R = cd^{1/2}$, given the probability measure P corresponding to normalized area measure. Let $(x_1, \dots, x_n)$, $n = 3d$, denote the coordinates of $\mathbb{R}^{3d}$. Then the problem is to find the normalized area of the subset of S determined by $a \leq x_n \leq b$. In other words, we look for the measure of the interval $[a, b]$ with respect to the measure $\mu_d := x_{n*}P$. (By spherical symmetry, μ_d does not depend on the coordinate chosen.)

Assume that $-R < a < b < R$. (Notice that μ_d is supported on $[-R, R]$, that is, the complement of this interval has measure 0.) A simple calculation shows that the probability is given by

$$\mu_d([a,b]) = \frac{\int_a^b [1 - u^2/R^2]^{(3d-3)/2}\, du}{\int_{-R}^R [1 - u^2/R^2]^{(3d-3)/2}\, du}.$$

Using $\lim_{n\to\infty}(1-\alpha/n)^n = e^{-\alpha}$ and $\int_{-\infty}^{\infty} e^{-\frac{mu^2}{2kT}}\, du = (\frac{m}{2\pi kT})^{1/2}$ we obtain

$$\lim_{d\to\infty} \mu_d([a,b]) = (\frac{m}{2\pi kT})^{1/2} \int_a^b e^{-\frac{mu^2}{2kT}}\, du.$$

Therefore, the components of the velocity of the gas molecules are normally distributed, with density

$$\rho(x) = (\frac{m}{2\pi kT})^{1/2} e^{-\frac{mx^2}{2kT}}.$$

It follows that the probability density for the velocity of individual molecules is given by the *Maxwell-Boltzmann* distribution

$$\rho(v) = (\frac{m}{2\pi kT})^{3/2} e^{-\frac{m|v|^2}{2kT}}.$$

This shows that v is a normal random variable with mean 0 and variance kT/m.

In order to understand the significance of the temperature parameter T, consider the following remark. Suppose that the insulated box contains two types of (point) particles that can be distinguished by their masses m_1 and m_2. The total number of molecules is d, a fraction f_1 comprised of molecules of mass m_1 and and a fraction $f_2 = 1 - f_1$ of molecules of mass m_2. The box is thermally insulated so that the total energy E is kept constant. Let E_1 and E_2 be the total (kinetic) energies of the gas components of type 1 and 2, respectively, so that $E_1 + E_2 = E$. We write, for $i = 1, 2$:

$$E_i = \frac{3}{2} k T_i f_i d.$$

Also introduce the parameter $T = f_1 T_1 + f_2 T_2$. This is chosen so that $E = (3/2)kTd$.

The (velocity part of the) phase space for the gas can be described by the ellipsoid

$$\{(v, w) \in \mathbb{R}^{3d_1} \times \mathbb{R}^{3d_2} \mid \frac{1}{2} m_1 |v|^2 + \frac{1}{2} m_2 |w|^2 = E\}.$$

As before, we assume that the distribution of velocities is uniform over the ellipsoid, that is, the probability P is proportional to the area measure and we ask for the probability distribution of T_1. A simple calculation (that uses the observation that the subset of the phase space corresponding to a given value of T_1 is the cartesian product of two spheres of radii $(2E_i/m_i)^{1/2}$, for $i = 1, 2$) shows that for large values of d (the total number of particles) the probability of $x := f_1 T_1 / T$, over the interval $[0, 1]$, has density

$$c[x^{f_1}(1-x)^{f_2}]^d.$$

Again for large values of d, this density has a sharp maximum at the point $x = f_1/(f_1 + f_2)$, which corresponds to the value

$$T_1 = T = T_2.$$

The conclusion is that, under the uniformity assumption made about P, the part of phase space most likely to be occupied by the gas corresponds to that for which the temperature parameters in both components of the gas mixture coincide.

This argument suggests that the probability density for the velocity of a particle of mass m immersed in a monoatomic gas at temperature T will approach in time the equilibrium value given by the Maxwell-Boltzmann distribution with mass m and temperature T.

We would like now to take a closer look at how the microscopic interactions between a particle of mass m suspended on a gas at temperature T of monoatomic molecules of much smaller mass, can lead to the probability density given in the previous paragraph. We call the larger particle a *Brownian particle*.

The Brownian particle is assumed to be under two kinds of forces. One is a frictional force due to viscosity, given by $-m\beta v$, where β is a constant and v is the particle's velocity. The second force is due to the combined effect of individual collisions with the surrounding molecules and has a highly fluctuating and chaotic behavior. We denote it by $mf(t)$. By Newton's second law of motion, $v(t)$ is described by the differential equation

$$\frac{dv}{dt} = -\beta v + f(t).$$

We regard $f(t)$ as a random variable, and expect that a solution of this equation (if one exists in some appropriate sense) will also be a random variable. If $\rho_{v_0}(v,t)$ denotes the density of the probability distribution of $v(t)$, (conditioned by $v(0) = v_0$) we expect, given the earlier discussion, that as t grows,

$$\rho_{v_0}(v,t) \to (\frac{m}{2\pi kT})^{3/2} e^{-\frac{m|v|^2}{2kT}}$$

and as t approaches 0, $\rho_{v_0}(v,t)$ approaches the Dirac delta function concentrated at v_0. The above equation is called the *Langevin equation*. Using the equilibrium value of the velocity distribution, we would like to determine the statistical properties of f.

The formal solution of the Langevin equation is

$$v(t) = v_0 e^{-\beta t} + e^{-\beta t} \int_0^t e^{\beta s} f(s)\ ds.$$

As the first term on the right-hand side of the equation goes to 0, the probability density of $\int_0^t e^{-\beta(t-s)} f(s)\ ds$ should approach, for large t, the Maxwell-Boltzmann equilibrium density of v.

Consider the Riemann sum approximation:

$$\int_0^t e^{-\beta(t-s)} f(s)\ ds \approx e^{-\beta t} \sum_i e^{\beta i \Delta t} f(i\Delta t)\ \Delta t.$$

The random variables $\Delta b_i := f(i\Delta t)\ \Delta t$ express the accelerations that the Brownian particle gains during the interval $[i\Delta t, (i+1)\Delta t]$. For large t, we have:

$$v \approx \sum_i e^{\beta i \Delta t - t} \Delta b_i.$$

Δb_i is assumed to be the sum of the accelerations due to a large number of independent collisions with the surrounding molecules, taking place during the interval $[i\Delta t, (i+1)\Delta t]$. It is thus natural to suppose that the Δb_i are independent random variables with the same kind of probability distribution as for the molecular velocities, that is, the Δb_i are assumed to be independent equally distributed normal random variables of 0 mean. To determine the variance $V(\Delta t)$ of Δb_i, we use that the limit (Maxwell-Boltzmann) distribution has variance $\frac{kT}{m}$, so that

$$E\left[|v|^2\right] \to \frac{kT}{m}.$$

On the other hand (using that the Δb_i are independent),

$$\begin{aligned} E\left[|v|^2\right] &\approx E\left[\sum_{i,j} e^{\beta(i\Delta t - t)+\beta(j\Delta t - t)}\langle \Delta b_i, \Delta b_j\rangle\right] \\ &= \sum_{i,j} e^{\beta(i\Delta t - t)+\beta(j\Delta t - t)} E[\langle \Delta b_i, \Delta b_j\rangle] \\ &= \sum_i e^{2\beta(i\Delta t - t)} E\left[|\Delta b_i|^2\right] \\ &= \sum_i e^{2\beta(i\Delta t - t)} V(\Delta t). \end{aligned}$$

In order for the Riemann sums to converge as $\Delta t \to 0$ it is now apparent that we must require $V(\Delta t)/\Delta t$ to have a nonzero finite limit. We call the limit σ^2. Therefore,

$$E\left[|v|^2\right] \approx \sum_i e^{2\beta(i\Delta t - t)} \sigma^2 \Delta t \approx \sigma^2 \int_0^t e^{-2\beta(t-s)}\, dt = \frac{\sigma^2}{2\beta}(1 - e^{-2\beta t})$$

and we obtain $\sigma^2 = 2\beta kT/m$.

The conclusion is that the "force" f that accounts for the Maxwell-Boltzmann distribution of the Brownian particle is expected to be $f(t)\Delta t = \Delta b$, where Δb is a normal random variable of mean 0 and variance $2\beta kT/m$.

4. The Volume Concentration Phenomenon

CHAPTER 2

Probability Spaces

1. The Language of Measure Theory

a. Measurable spaces. Let Ω be an arbitrary set. The complement of a subset $A \subset \Omega$ will be denoted alternatively by $\Omega - A$ or A^c. A *σ-algebra* $\mathcal{F}$ on Ω is a non-empty family of subsets of Ω satisfying the following properties:

(1) If a set A belongs to $\mathcal{F}$, then its complement $\Omega - A$ also belongs to $\mathcal{F}$. In particular, $\Omega \in \mathcal{F}$.
(2) If $A_i \in \mathcal{F}$ for each $i = 1, 2, 3, \ldots$, then the union

$$A = \bigcup_{i=1}^{\infty} A_i$$

also belongs to $\mathcal{F}$.

A family $\mathcal{F}$ satisfying these properties is called a *σ-algebra* on Ω. Elements of $\mathcal{F}$ are called *measurable sets.* Note that Ω and the empty set $\emptyset$ must be contained in $\mathcal{F}$, since $\mathcal{F}$ is nonempty and $\Omega^c = \emptyset = A \cap A^c$ for any $A \in \mathcal{F}$. If more than one σ-algebra is in view and confusion could arise, these sets will be called $\mathcal{F}$-measurable sets. The pair $(\Omega, \mathcal{F})$ is called a *measurable space.*

Starting from a collection $\mathcal{A}$ of subsets of Ω, there exists a smallest σ-algebra containing $\mathcal{A}$, called the σ-algebra *generated* by $\mathcal{A}$, defined by the intersection of all σ-algebras of Ω containing $\mathcal{A}$. (There always exists a σ-algebra containing $\mathcal{A}$, namely, the family of all subsets of Ω.) If Ω is a topological space and $\mathcal{A}$ is the collection of all open sets in Ω, the σ-algebra $\mathcal{B}$ generated by $\mathcal{A}$ is called the *Borel σ-algebra* on Ω. The elements of $\mathcal{B}$ are called *Borel sets.* Open sets, closed sets, countable unions or intersections of open or closed sets are also Borel.

b. Measurable Functions, or Random Variables. Given two measurable spaces $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$, a function $f : \Omega \to \Omega'$ is said to be a *measurable function* or, if we need to be explicit, a $\mathcal{F}/\mathcal{F}'$-measurable function, if $f^{-1}(B) \in \mathcal{F}$ for each $B \in \mathcal{F}'$. Measurable functions will often be called *random variables.* If Ω and Ω' are both topological spaces with the respective Borel σ-algebras, then continuous functions from Ω to Ω' are clearly measurable.

Let $f : \Omega \to \Omega'$ be any function from a set Ω to a measurable space $(\Omega', \mathcal{F}')$, the σ-algebra $\mathcal{F}_f$ *generated by* f is the smallest σ-algebra on Ω containing all $f^{-1}(A)$ for $A \in \mathcal{F}'$, which is precisely the set of all $f^{-1}(A)$ for $A \in \mathcal{F}'$. It will also be denoted by $f^{-1}(\mathcal{F}')$. As a particular case, if Ω is a subset (not necessarily measurable) of Ω' and f is the inclusion map, then $f^{-1}(\mathcal{F}')$ is the collection of all $\Omega \cap A$, $A \in \mathcal{F}'$.

The product of an arbitrary collection of measurable spaces $(\Omega_\alpha, \mathcal{F}_\alpha)$, $\alpha \in I$, is a measurable space $(\Omega, \mathcal{F})$, where $\Omega = \Pi_\alpha \Omega_\alpha$ is the Cartesian product of the Ω_α

and $\mathcal{F}$ is the σ-algebra generated by the sets $\pi_\alpha^{-1}(A)$, where $A \in \mathcal{F}_\alpha$, $\alpha \in I$, and $\pi_\alpha : \Omega \to \Omega_\alpha$ is the natural projection.

A *partition* of Ω is a (finite or countable) collection of disjoint subsets A_i, $i = 1, 2, \ldots$, such that $\cup_i A_i = \Omega$. A partition is said to be $\mathcal{F}$-measurable if each A_i belongs to $\mathcal{F}$. An $\mathcal{F}$-measurable function f is said to be a *simple function* if it is constant over each element of a finite $\mathcal{F}$-measurable partition.

PROPOSITION 2.1.1. A function $g : \Omega \to \mathbb{R}$ is measurable with respect to a σ-algebra $\mathcal{F}$ on Ω (and the Borel σ-algebra on $\mathbb{R}$) if and only if there exist ($\mathcal{F}$-measurable) simple functions g_n such that

$$\lim_{n\to\infty} g_n(\omega) = g(\omega)$$

for each $\omega \in \Omega$.

PROOF. Define for any pair of integers k, n, $n \geq 0$, a measurable set

$$A_{k,n} := f^{-1}([\frac{k}{2^n}, \frac{k+1}{2^n})).$$

Let g_n be the function that assumes the value $g_n(\omega) = k/2^n$ for $\omega \in A_{k,n}$ and $g_n(\omega) = (k+1)/2^n$ if $k = -1, \ldots, -n2^n$. Then g is the pointwise limit of the g_n, which are all measurable. Conversely, if g_n are measurable and g is the pointwise limit of the g_n, then g is $\mathcal{F}$-measurable. This can be seen by writing

$$g^{-1}((t,\infty)) = \bigcup_{k=1}^{\infty} \bigcup_{n=i}^{\infty} \bigcap_{j=n}^{\infty} \{\omega \in \Omega \mid h_j(\omega) > t + \frac{1}{k}\}.$$

□

THEOREM 2.1.2 (Doob-Dynkin lemma). Let $f : \Omega \to X$ be a measurable function between measurable spaces $(\Omega, \mathcal{F})$ and $(X, \mathcal{A})$. Then, a function $g : \Omega \to \mathbb{R}^n$ is measurable with respect to the σ-algebra $f^{-1}(\mathcal{A})$ if and only if there exists a measurable function $f : X \to \mathbb{R}^n$ such that $g = h \circ f$.

PROOF. It clearly suffices to consider the case $n = 1$. We first assume that g is a simple function. Let $\{A_i \mid 1 \leq i \leq m\}$ be a measurable partition of Ω and suppose that $g(\omega) = a_i$ for $\omega \in A_i$. Since A_i is $f^{-1}(\mathcal{A})$-measurable, there exists by definition a $B_i \in \mathcal{A}$ such that $A_i = f^{-1}(B_i)$, for each i. Define an $\mathcal{A}$-measurable partition $\{C_i \mid 1 \leq i \leq m\}$ of X by setting $C_1 := B_1$ and

$$C_i := B_i - B_1 \cup B_2 \cup \cdots \cup B_{i-1}$$

for each i, $2 \leq i \leq n$. It is immediate that the C_i are indeed disjoint and form a partition of X. Furthermore, $f^{-1}(C_i) = f^{-1}(B_i) = A_i$ and $h \circ f = g$ as required.

Let now g be $f^{-1}(\mathcal{A})$-measurable but not necessarily simple. Let g_n be $f^{-1}(\mathcal{A})$-measurable simple functions such that g_n converges pointwise to g. We have seen that for each n there is an $\mathcal{A}$-measurable function $h_n : X \to \mathbb{R}$ such that $g_n = h_n \circ f$. Let X_0 denote the subset of X consisting of x such that $h_n(x)$ is a Cauchy sequence. Notice that X_0 is $\mathcal{A}$-measurable since

$$X_0 = \bigcap_{L\in\mathbb{N}} \bigcup_{N\in\mathbb{N}} \bigcap_{n,m\geq N} H_{n,m}^{-1}(-\frac{1}{L}, \frac{1}{L})$$

where $H_{n,m} := h_n - h_m$.

Now define $h(x) = \lim_{n\to\infty} h_n(x)$ if $x \in X_0$ and 0 otherwise. Then h is $\mathcal{A}$-measurable (since it is a pointwise limit of measurable functions) and we have $g = h \circ f$, as needed. □

c. Spaces of Paths. The previous theorem, and much of the measure theory of $\mathbb{R}^n$, still holds if we replace $\mathbb{R}^n$ with a general Polish space, that is, a separable topological space that admits a complete metric. This is because a Polish space is measurably isomorphic to a measurable subset of $\mathbb{R}$. A proof of this fact will be given later when we discuss the notion of measurable isomorphism.

An important example of a Polish space is as follows. Let $\Omega = C([0,\infty), \mathbb{R}^n)$ denote the space of continuous paths in $\mathbb{R}^n$, that is, of continuous functions $f : [0,\infty) \to \mathbb{R}^n$. Give Ω the *topology of uniform convergence on compact intervals.* In this topology a neighborhood basis of an element $\omega \in \Omega$ consists of sets

$$U_\omega(\epsilon, T) = \{\eta \in \Omega : \sup_{t\in[0,T]} |\eta(t) - \omega(t)| < \epsilon\}$$

for $T > 0, \epsilon > 0$.

See §3.1 of Daniel W. Stroock, *Probability Theory – an analytic view*, Cambridge U. Press, 1993

The resulting topology is metrizable. In fact, a metric that generates the topology of uniform convergence on compact intervals can be defined by the following expression:

$$d(\omega_1, \omega_2) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{\sup_{t\in[0,n]} |\omega_1(t) - \omega_2(t)|}{1 + \sup_{t\in[0,n]} |\omega_1(t) - \omega_2(t)|}$$

for $\omega_1, \omega_2 \in \Omega$. It can be shown that (Ω, d) is a complete separable metric space.

d. Probability. A *probability measure* P on a measurable space $(\Omega, \mathcal{F})$ is a real valued function P on $\mathcal{F}$ that satisfies the following properties (in words: a countably additive nonnegative measure with total mass 1):

(1) $0 \le P(A) \le 1$ for each $A \in \mathcal{F}$
(2) If $A_1, A_2, \ldots$ is a finite or countably infinite sequence of disjoint elements of $\mathcal{F}$ (that is, $A_i \cap A_j = \varnothing$ for all i and j distinct), then

$$P(\bigcup_i A_i) = \sum_i P(A_i)$$

(3) $P(\Omega) = 1$.

An immediate consequence of the axioms is that $P(\varnothing) = 0$.

The fundamental notion in probability theory is a *probability space*, which formalizes the idea of an experiment with uncertain outcome.

Definition 2.1.3. A probability space is a triple $(\Omega, \mathcal{F}, P)$ where $(\Omega, \mathcal{F})$ is a measurable space and P is a probability measure. The set Ω represents the collection of all possible *outcomes* of an experiment, the elements of $\mathcal{F}$ are the possible *events* - an event $A \in \mathcal{F}$ being *realized* if an outcome $\omega \in \Omega$ belongs to A, and for each $A \in \mathcal{F}$, $P(A)$ is the probability of event A.

A property or condition on the sample space Ω is said to hold *almost surely* (or *almost everywhere*) if the subset of $\omega \in \Omega$ satisfying the property has measure 1.

e. Completion of a σ-algebra. It is often useful to suppose that the measure space is complete, in the sense that if A is a measurable set, then any A' that differs from A by a set (that is contained in a measurable set) of measure 0 is also measurable. More precisely, given a measure space $(\Omega, \mathcal{F}, P)$, define a family $\mathcal{F}^P$ consisting of $A \subset \Omega$ for which one can find $B_1, B_2 \in \mathcal{F}$ such that

$$P(B_1) = P(B_2), \quad B_1 \subset A \subset B_2.$$

Define $P^*(A) := P(B_1)$. Then $\mathcal{F}^P$ is a σ-algebra and P^* is a probability on $\mathcal{F}^P$. The triple $(\Omega, \mathcal{F}^P, P^*)$ is called the *completion* of $(\Omega, \mathcal{F}, P)$.

A probability P on $(\Omega, \mathcal{F})$ is said to be *complete* if $\mathcal{F} = \mathcal{F}^P$. In this case, the measure space is called a *complete probability space*.

f. Push-forward of a Measure. Let $(\Omega', \mathcal{F}')$ be a measurable space and $X : \Omega \to \Omega'$ a measurable function. Then X carries P to a probability measure $P_X := X_* P$ on Ω' defined by

$$(X_* P)(A) := P(X^{-1}(A))$$

for each $A \in \Omega'$. P_X will sometimes be called the *law* (or *distribution*) of the random variable X. Two random variables, possibly defined on different probability spaces (but with same target space) are said to *agree in law* if their distributions coincide.

g. Integration. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $f : \Omega \to \mathbb{R}$ a nonnegative measurable function.

Suppose first that g has a *discrete distribution*, that is, there exists a countable set $K = \{a_1, a_2, \dots\}$ of values of f such that $P(\{\omega \in \Omega : f(\omega) \in K\}) = 1$. Define

$$E[f] = \sum_{i=1}^{\infty} a_i P(\{\omega \in \Omega : f(\omega) = a_i\})$$

provided the sum is absolutely convergent.

For a general f, not necessarily with discrete distribution, E can be defined by approximation as follows. Denote by $A_{k,n} \in \mathcal{F}$ the set where $\frac{k}{n} \leq f \leq \frac{k+1}{n}$ and define $\underline{f_n} = \frac{k}{n} \chi_{A_{k,n}}$ and $\overline{f_n} = \frac{k+1}{n} \chi_{A_{k,n}}$. Observe that $\underline{f_n} \leq f \leq \overline{f_n}$ and that $|\overline{f_n} - \underline{f_n}| \leq \frac{1}{n}$. This allows to define

$$E[f] = \lim_{n\to\infty} E[\underline{f_n}] = \lim_{n\to\infty} E[\overline{f_n}]$$

provided $E[\overline{f_n}]$ exists for some n. If it does, we say that f is integrable and that $E[f]$ is the *expectation* of f, or the integral of f. We also write

$$E[f] = \int_\Omega f dP.$$

We will often denote the integral of f on Ω by

$$\int_\Omega f(\omega)\ dP(\omega).$$

When f is not necessarily nonnegative and f^+, f^- are the positive and negative parts of f (both nonnegative functions such that $f = f^+ - f^-$), then

$$\int_\Omega f(\omega)\ dP(\omega) := \int_\Omega f^+(\omega)\ dP(\omega) - \int_\Omega f^-(\omega)\ dP(\omega)$$

assuming that at least one of the two integrals is finite.

The law X_*P of a random variable X defined on a probability space $(\Omega, \mathcal{F}, P)$ is characterized by the property

$$\int_\Omega f \circ X \; dP = \int_{\Omega'} f \; d(X_*P)$$

for every measurable function $f : \Omega' \to \mathbb{R}$. (When f is the characteristic function χ_A of a set A, then $\chi_A \circ X$ is the characteristic set of $X^{-1}(A)$ and the above equality of integrals reduces to the definition of X_*P. The identity can be proved by approximating f by simple functions.)

h. Expectation, Variance, and $L^2(\Omega, \mathcal{F}, P)$. If X takes values in $\mathbb{R}^n$ and is *integrable*, that is, if

$$\int_\Omega |X(\omega)| \; dP(\omega) < \infty$$

then the *expectation* of X is defined by

$$E[X] := \int_\Omega X(\omega) \; dP(\omega)$$

Notice that if $f : \mathbb{R}^n \to \mathbb{R}$ is a measurable function, then

$$E[f \circ X] = \int_{\mathbb{R}^n} f(x) \; d(P_X)(x).$$

(This is immediate if $X = \sum a_i \chi_{A_i}$ is a simple function. The general case follows by taking limits.) In particular,

$$E[X] = \int_{\mathbb{R}^n} x \; d(P_X)(x).$$

Let $m = E[X]$ denote the mean, or expectation, of the random variable X. The *variance* of X is defined by

$$\mathrm{var}[X] := E[(X - m)^2] = \int_{\mathbb{R}^n} |x - m|^2 \; d(P_X)(x).$$

We say that a random variable $X : \Omega \to \mathbb{C}$ is *square integrable* if

$$\|X\|^2 := \int_\Omega |X|^2 dP < \infty.$$

The space of square integrable random variables will be denoted by $L^2(\Omega, \mathcal{F}, P)$. It is a Hilbert space. The norm $\|X\|^2$ comes from the inner product

$$\langle X, Y \rangle := \int_\Omega X\bar{Y} dP = E[X\bar{Y}],$$

where $\bar{Y}$ denotes the complex conjugate of Y.

The cosine of the angle between $X - E[X]$ and $Y - E[Y]$ is called the *correlation coefficient* of X and Y. The two random variables are said to be *uncorrelated* if $X \perp Y$; equivalently, their correlation coefficient is 0. It is clear that if X and Y are independent, then they are also uncorrelated.

The *covariant matrix* of a family of random variables $X_1, X_2, \ldots, X_n$ is the matrix whose (i, j)-entry is

$$\mathrm{cov}(X_i, X_j) := \langle X_i - E[X_i], X_j - E[X_j] \rangle.$$

i. Types of Convergence. Let $X, X_k,\ k = 1, 2, \ldots$ be random variables on a probability space $(\Omega, \mathcal{F}, P)$, taking values in $\mathbb{R}$.

We will consider the following types of convergence:

1. *Convergence Almost Everywhere.* X_k converges almost everywhere, or almost surely, to X if

$$\lim_{k\to\infty} X_k(\omega) = X(\omega)$$

for all $\omega \in \Omega$ with the possible exception of the ω in a subset of P-measure 0.

2. *Convergence in Probability.* X_k converges to X *in probability* if, for every $\epsilon > 0$,

$$P(\{\omega \in \Omega : |X_k(\omega) - X(\omega)| > \epsilon\}) \to 0.$$

3. *Convergence in L^p.* X_k converges to X in L^p-norm if

$$E[|X_k - X|^p] \to 0.$$

4. *Convergence in Law.* X_k converges to X in law if $\mu_k := (X_k)_* P$ converges to $\mu := X_* P$. This means that for every bounded uniformly continuous function $f : \mathbb{R}^n \to \mathbb{R}$,

$$\int_{\mathbb{R}^n} f d\mu_k \to \int_{\mathbb{R}^n} f d\mu.$$

j. Absolute Continuity and the Radon-Nikodym Derivative. Given two measures P, Q on a measurable space $(\Omega, \mathcal{F})$, we say that P is *absolutely continuous* with respect to Q if $P(A) = 0$ for all $A \in \mathcal{F}$ such that $Q(A) = 0$.

THEOREM 2.1.4 (Radon-Nikodym). If Q and P are probability measures on a measurable space $(\Omega, \mathcal{F})$ such that P is absolutely continuous with respect to Q, then there exists a unique (Q-a.e.) nonnegative Q-integrable measurable function h such that

$$P(A) = \int_A h\ dQ$$

for all $A \in \mathcal{F}$. A function $g : \Omega \to \mathbb{C}$ is P-integrable if and only if gh is Q-integrable.

The function h obtained in the theorem is usually denoted $\frac{dP}{dQ}$, and is called the *Radon-Nikodym derivative* of P with respect to Q.

Let $C(\Omega)$ be the linear space of (real valued) continuous functions on Ω. We define for $f \in C(\Omega)$

$$\|f\| := \sup\{|f(\omega)| : \omega \in \Omega\},$$

which is a finite norm on elements of $C(\Omega)$. With this norm, $C(\Omega)$ is a (real) Banach space. Its dual space, $C(\Omega)^*$, consisting of bounded linear functionals on $C(\Omega)$, is given the *weak* topology*, defined as follows: a sequence $\Lambda_n \in C(\Omega)^*$ converges to $\Lambda \in C(\Omega)^*$ if for all $f \in C(\Omega)$ the sequence of numbers $\Lambda_n(f)$ converges to $\Lambda(f)$.

THEOREM 2.1.5 (Riesz representation theorem). Let Ω be a compact Hausdorff space and Λ a positive linear functional on $C(\Omega)$. Then there exists a regular Borel measure P on Ω such that

$$\Lambda(f) = \int_\Omega f\ dP$$

for all $f \in C(\Omega)$. If $\Lambda(1) = 1$, P is a probability measure.

By the Riesz representation theorem, the space of Borel probability measures on Ω can be viewed as the set of bounded positive linear functionals Λ on $C(\Omega)$ such that $\Lambda(1) = 1$.

k. Measurable isomorphism. Two measurable spaces $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ are said to be *measurably isomorphic* if there is a bijection Φ between them such that Φ and Φ^{-1} are both measurable. If P and P' are probability measures on Ω and Ω' such that $\Phi_* P = P'$, then the probability spaces $(\Omega, \mathcal{F}, P)$ and $(\Omega', \mathcal{F}', P')$ are, for most probabilistic purposes, indistinguishable.

As an example, consider the set $\Omega = [0,1)^{\mathbb{N}}$ of all sequences of nonnegative real numbers less than 1. We give it the product measurable structure. We show that Ω is measurably isomorphic to the interval $[0,1)$.

An isomorphism $\Phi : [0,1) \to [0,1)^{\mathbb{N}}$ is constructed as follows. Each $x \in [0,1)$ has a unique binary expansion

$$x = 0.x_1 x_2 x_3 \cdots$$

containing infinitely many zeros. Now write $\Phi(x) = (\Phi_1(x), \Phi_2(x), \dots)$, where $\Phi_i(x) = 0.y_1 y_2 y_3 \cdots$ and $(y_1, y_2, \cdots)$ is the i-th column of the following scheme:

$$\begin{array}{ccccc} x_1 & & & & \\ x_2 & x_3 & & & \\ x_4 & x_5 & x_6 & & \\ x_7 & x_8 & x_9 & x_{10} & \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

Thus,

$$\begin{aligned} \Phi_1(x) &= 0.x_1 x_2 x_4 \cdots \\ \Phi_2(x) &= 0.x_3 x_5 x_8 \cdots \\ \Phi_3(x) &= 0.x_6 x_9 x_{13} \cdots \\ &\cdots \end{aligned}$$

It is clear that Φ is bijective and that its inverse is the function

$$\Phi^{-1} : [0,1)^{\mathbb{N}} \to [0,1)$$

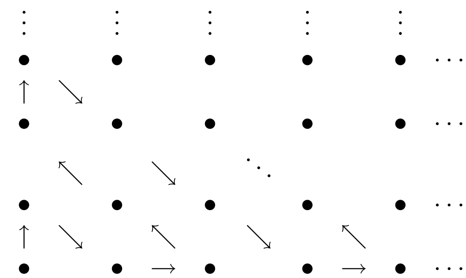
that sends $(x^1, x^2, \dots)$ to x as follows. Write $x^s = 0.x^s_1 x^s_2 \cdots$ – the unique binary expansion with infinitely many zeros. Then

$$x = \Phi^{-1}(x^1, x^2, \cdots) = 0.y_1 y_2 y_3 \cdots$$

where

$$\begin{pmatrix} y_1 & & & \\ y_2 & y_3 & & \\ y_4 & y_5 & y_6 & \\ y_7 & y_8 & y_9 & y_{10} \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} = \begin{pmatrix} x^1_1 & & & \\ x^1_2 & x^2_1 & & \\ x^1_3 & x^2_2 & x^3_1 & \\ x^1_4 & x^2_3 & x^3_2 & x^4_1 \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

To prove that Φ and its inverse are measurable, it will be helpful to define the bijection $a : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ such that $y_{a(i,j)} = x^i_j$. Notice that a corresponds to the enumeration of $\mathbb{N} \times \mathbb{N}$ according to the following scheme (the dots are the lattice points of $\mathbb{N} \times \mathbb{N}$):



Define $I_0 = [0, 1/2)$ and $I_1 = [1/2, 1)$, and the measurable (piecewise continuous) bijection $T : [0, 1) \to [0, 1)$ by

$$T(x) = \begin{cases} 2x & \text{if } x \in I_0 \\ 2x - 1 & \text{if } x \in I_1. \end{cases}$$

The ith binary digit of x is given by the measurable function

$$d_i(x) = \chi_{I_1}(T^i(x))$$

where χ_{I_1} is the characteristic function of the interval I_1 and $T^i = T \circ \cdots \circ T$, is the ith iterate of T. For any sequence of integers $S = (n_1, n_2, \dots)$ the function

$$\Phi_S(x) := \sum_{i=1}^{\infty} d_{n_i}(x) 2^{-i}$$

is, therefore, measurable. Since each Φ_i is of this form, Φ is measurable.

Denoting by $\pi_i : [0, 1)^{\mathbb{N}} \to [0, 1)$ the natural projection onto the ith factor, we can write:

$$\Phi^{-1}(\omega) = \sum_{i,j} d_i(\pi_j(\omega)) 2^{-a(i,j)}$$

which is also measurable.

Let Ω be a separable topological space that admits a complete metric, endowed with the Borel σ-algebra. There is no loss of generality in assuming that the metric d takes values in $[0, 1)$ since, otherwise, we could simply replace d with the new metric $d(\cdot, \cdot)/[1 + d(\cdot, \cdot)]$. Being separable means that Ω admits a countable dense subset $\{\omega_1, \omega_2, \dots\}$. A separable topological space that admits a complete metric is called a *Polish space*. Most spaces of interest in probability theory are Polish spaces. For example, it can be shown that if M is a manifold (say, $X = \mathbb{R}^n$) then the space $C([0, \infty), M)$ of all continuous functions from $[0, \infty)$ into M is a Polish space.

Proposition 2.1.6. A Polish space Ω is homeomorphic to a countable intersection of open sets in $[0, 1]^{\mathbb{N}}$. It is also measurably isomorphic to a measurable subset of $[0, 1]$.

Proof. Define $\Phi : \Omega \to [0, 1]^{\mathbb{N}}$ by

$$\Phi(\omega) := (d(\omega, \omega_1), d(\omega, \omega_2), \dots)$$

where $\{\omega_1, \omega_2, \dots\}$ is a countable dense subset of Ω. Since each component of Φ is continuous, Φ is also continuous (for the product topology on $[0, 1]^{\mathbb{N}}$). Φ is clearly one-to-one.

We claim that Φ is a homeomorphism between Ω and $\Phi(\Omega)$. To see that Φ^{-1} is continuous we need to show that if a sequence $\eta_i \in \Omega$ does not converge to some η, then $\Phi(\eta_i)$ does not converge to $\Phi(\eta)$. But if η_i does not converge to η, $d(\eta_i, \eta) > \epsilon$ for some positive ϵ and infinitely many i, and by the triangle inequality, if l is such that $d(\eta, \omega_l) < \epsilon/2$, then $d(\eta_i, \omega_l) > \epsilon/2$ for infinitely many i. This means that the value of the lth component of Φ on η_i does not converge to its value on η. In particular, $\Phi(\eta_i)$ does not converge to $\Phi(\eta)$.

For each $\omega \in \Omega$, let U be an open neighborhood of $\Phi(\omega)$. We choose U_ω so that U_ω and $\Phi^{-1}(U_\omega)$ have diameter less than $1/n$, with respect to a choice of metric on $[0,1)^{\mathbb{N}}$ and d, respectively. Then $\Phi(\Omega)$ is contained in the union, U_n, of the U_ω for each n. Therefore, $\Phi(\Omega)$ is contained in the intersection $G = \cap U_n$. The first part of the proposition will be proved if we show that $G = \Phi(\Omega)$.

Let $x \in G$. For each n there exists an open ball B_n containing x, having a center at $\Phi(\eta_n)$, for some $\eta_n \in \Omega$, and of the diameters of both B_n and $C_n := \Phi^{-1}(B_n)$ are than $1/n$. In particular, x is in the closure of $\Phi(\Omega)$, since $\Phi(\eta_n)$ must converge to x. Therefore, for each n_1 and n_2 the intersection $B_{n_1} \cap B_{n_2}$ contains some $x' \in \Phi(\Omega)$. The point $\omega' := \Phi^{-1}(x')$ belongs to $C_{n_1} \cap C_{n_2}$, and it follows (by the triangle inequality) that $d(\eta_{n_1}, \eta_{n_2}) \leq (1/n_1) + (1/n_2)$. Therefore, the η_n forms a Cauchy sequence. Since Ω is complete, η_n converges to some ω. By continuity of Φ, we have $\Phi(\omega) = x$, so that $x \in G$ is in the image of Φ, as claimed. The remaining claim is a consequence of the measurable isomorphism between $[0,1)$ and $[0,1)^{\mathbb{N}}$ obtained earlier. □

We define the *compactification* of Ω by the closure of $\Phi(\Omega)$ in the compact space $[0,1]^{\mathbb{N}}$.

l. Kolmogorov's extension theorem. Let $(\Omega, \mathcal{F}, P)$ be a probability space and X_α, $\alpha \in I$, be a family of random variables,

$$X_\alpha : \Omega \to \Omega_\alpha$$

where, for each $\alpha \in I$, Ω_α is a Polish space. Let I_1, I_2 be finite subsets of I such that $I_1 \subset I_2$. We obtain projections

$$\Omega \to \prod_{\alpha \in I} \Omega_\alpha \to \prod_{\alpha \in I_2} \Omega_\alpha \to \prod_{\alpha \in I_1} \Omega_\alpha.$$

Under these projections P is sent to probability measures P_I, P_{I_2}, P_{I_1}. Writing $I_1 = \{\alpha_1, \dots, \alpha_k\}$, then P_{I_1} is simply the law of $(X_{\alpha_1}, \dots, X_{\alpha_l})$.

LEMMA 2.1.7. The σ-algebra generated by $\{X_\alpha | \alpha \in I\}$ coincides with the union of the σ-algebras generated by the $\{X_\alpha | \alpha \in J\}$, where J runs over all the countable subsets $J \subset I$.

PROOF. The σ-algebra generated by $\{X_\alpha | \alpha \in I\}$ clearly contains the union of the σ-algebras $\sigma(J)$ generated by $\{X_\alpha | \alpha \in J\}$ for each J. To prove equality, it suffices to show that this union is already a σ-algebra. If A belongs to the union $\cup_J \sigma(J)$, it must lie in some $\sigma(J)$, hence its complement (which lies in $\sigma(J)$) also belongs to the union. If $A_1, A_2, \dots$ is a countable sequence of sets in $\cup_J \sigma(J)$, then $A_i \in \sigma(J_i)$ for some J_i and each i. But $J := \cup_i J_i$ is countable, being a countable union of countable sets, and $A_i \in \sigma(J)$ for each i. Therefore, $\cup_i A_i$ also belongs to $\cup_J \sigma(J)$. Since the empty set is obviously in the union, all the conditions for a σ-algebra are satisfied. □

The lemma implies that any measurable set relative to the σ-algebra generated by an infinite family of random variables must be describable in terms of only countable many random variables in the family. Thus, if $I = [0,1]$ and $S_t = \mathbb{R}$, the set of ω in Ω such that $t \mapsto X_t(\omega)$ is continuous on $[0,1]$ may fail to be measurable.

THEOREM 2.1.8 (Kolmogorov's extension). Let S_α, $\alpha \in I$, be Polish spaces and for each finite subset $J \subset I$, let P_J be a probability measure on $S_J = \prod_J S_\alpha$, so that the family of probabilities is consistent. Then there exists a unique probability measure P on $S_I = \prod_I S_\alpha$ such that P_J is the image of P under the projection $S_I \to S_J$, for each J.

PROOF. We claim that it suffices to show the theorem for a countable set I. In fact, according to the lemma, If A is a measurable subset in the uncountable product $\prod_I S_\alpha$, then there must be a countable J such that A is measurable with respect to the σ-algebra $\sigma(J)$ generated by the projections π_α, $\alpha \in J$. Therefore, by the Doob-Dynkin lemma (applied to the characteristic function of A) the set A must be the inverse image of a unique measurable subset A' of $\prod_J S_\alpha$. Define $P(A) = P_J(A')$. If J' is another countable subset of I such that $A \in \sigma(J')$, then the uniqueness of the Kolmogorov's extension (for countable index sets) implies that P_J and $P_{J'}$ are both images of $P_{J\cup J'}$ under the respective projections, so that the definition of P does not depend on the choice of J. It is also easy to check that P satisfies the definition of a probability measure.

We now suppose that I is countable. The next claim is that it suffices to assume that S_α is a compact topological space for each α. In fact, suppose that the theorem is already proved for compact S_α. Recall that we can regard a Polish space S_α as a measurable subset of the compact space $[0,1]^\infty$, so that S_α can be compactified by taking its closure $\overline{S}_\alpha$ in $[0,1]^\infty$. Define $\overline{S}_J = \prod_J \overline{S}_\alpha$ and let $\bar{P}_J$ be the probability on $\overline{S}_J$ that extends P_J by assigning probability 0 to the complement of S_J in its closure. The family $\{\bar{P}_J\}$ is easily seen to be consistent, so that there exists a unique probability measure $\bar{P}$ on $\prod_I \overline{S}_\alpha$ that projects to the $\bar{P}_J$ under the natural projections $\overline{S}_I \to \overline{S}_J$, for each finite $J \subset I$. Since $\bar{P}_J(\overline{S}_J - S_J) = 0$ for each finite J, it follows that $\bar{P}_I(\overline{S}_I - S_I) = 0$. We claim that $P_I(\overline{S}_I - S_I) = 0$. In fact, the set $\overline{S}_I - S_I$ is the union of the (countably many) $R_\beta = \prod_I L_\alpha$, where $L_\alpha = \overline{S}_\alpha$ if $\alpha \neq \beta$ and $L_\beta = \overline{S}_\beta - S_\beta$. But $P_I(R_\beta) = P_\beta(\overline{S}_\beta - S_\beta) = 0$, so that $P_I(\overline{S}_I - S_I) = 0$. Therefore, $\bar{P}_I$ restricts to a probability measure P_I on S_I that satisfies the desired properties. The measure is easily seen to be unique.

It remains to prove the theorem for a countable I and compact S_α. Let $C(S_I)$ denote the space of continuous real valued functions on S_I. For each $J \subset I$ there is an inclusion $C(S_J) \to C(S_I)$ given by $g \mapsto g \circ \pi_J$, where $\pi_J : S_I \to S_J$ is the projection. Denote by D the union of the $C(S_J)$ (regarded as subsets of $C(S_I)$) for all finite J. It is clear that D is an algebra of functions (that is, given f and g in D, and any real λ, then $f + g$, fg and λf are all contained in D) and that it separates points in S_I. By the Stone-Weierstrass theorem, D is dense in $C(S_I)$.

Let $F : D \to \mathbb{R}$ be the linear functional on D defined as follows. If $f \in D$, then $f = g \circ \pi_J$ for some $g \in C(S_J)$ and for some finite J, set $F(f) := \int_{S_J} g \, dP_J$. Then F is a bounded linear functional such that $F(f) \geq 0$ for $f \geq 0$ and $F(f) = 1$ if $f \equiv 1$. By the Hahn-Banach theorem, F extends to a bounded linear functional $\bar{F}$ on $C(S_I)$. The extension also satisfies $\bar{F}(f) = 1$ for $f \equiv 1$ and $\bar{F}(f) \geq 0$ whenever $f \geq 0$.

By the Riesz representation theorem, there exists a probability measure P on S_I such that $\bar{F}(f) = \int_{S_I} f \, dP$ for all $f \in C(S_I)$. P is unique since a probability measure on a compact space is determined by its values on continuous functions, and D is dense in the space of continuous functions. It is immediate to check that P satisfies the required properties. □

2. Independence and Conditional Expectation

a. Independence. Probability theory begins to assume a separate identity from general measure theory when the notions of independence and conditional expectation are introduced.

Let $(\Omega, \mathcal{F}, P)$ be a probability space. A collection $A_1, A_2, \dots, A_k$ of sets in $\mathcal{F}$ is said to be *independent* if

$$P(A_1 \cap \cdots \cap A_k) = P(A_1) \cdots P(A_k).$$

A collection of σ-algebras $\mathcal{F}_1, \dots, \mathcal{F}_k$ contained in $\mathcal{F}$ is said to be independent if for any choice $A_i \in \mathcal{F}_i$, the sets $A_1, \dots, A_k$ are independent. A collection of measurable functions $X_i : \Omega \to \Omega_i$, $i = 1, \dots, k$, taking values into measurable spaces $(\Omega_i, \mathcal{F}_i)$, is independent if the σ-algebras $X_i^{-1}(\mathcal{F}_i)$ generated by X_i are independent.

Proposition 2.2.1. Two integrable random variables $X, Y : \Omega \to \mathbb{R}$ are independent if and only if for all bounded (Borel) measurable functions f, g on $\mathbb{R}$

$$E[(f \circ X)(g \circ Y)] = E[f \circ X] E[g \circ Y].$$

Proof. If X and Y are independent, the σ-algebras $\mathcal{F}_X$ and $\mathcal{F}_Y$ generated by X and Y are, by definition, independent. Let X_n (respectively, Y_n) be the $\mathcal{F}_X$-measurable (respectively, $\mathcal{F}_Y$-measurable) bounded simple function such that $E[X_n]$ converges to $E[X]$ ($E[Y_n]$ converges to $E[Y]$), which was constructed in the section where the expectation (Lebesgue integral) is defined. Notice that X_n and Y_n are still independent.

Since the sequence $E[(f \circ X_n)(g \circ Y_n)]$ converges to $E[(f \circ X)(g \circ Y])$, it suffices to prove the proposition for simple functions.

Write $X_n = \sum_{i=1}^{l} a_i \chi_{A_i}, Y_n = \sum_{j=1}^{k} b_j \chi_{B_j}$, and observe that

$$f \circ X_n = \sum_{i=1}^{l} f(a_i) \chi_{A_i}, \quad g \circ Y_n = \sum_{j=1}^{k} g(b_j) \chi_{B_j}.$$

Then

$$\begin{aligned}
E[(f\circ X_n)(g\circ Y_n)] &= \sum_{i,j} f(a_i)g(b_j)E[\chi_{A_i}\chi_{B_j}] \\
&= \sum_{i,j} f(a_i)g(b_j)E[\chi_{A_i\cap B_j}] \\
&= \sum_{i,j} f(a_i)g(b_j)P(A_i\cap B_j) \\
&= \sum_{i,j} f(a_i)g(b_j)P(A_i)P(B_j) \\
&= (\sum_i f(a_i)P(A_i))(\sum_j g(b_j)P(B_j)) \\
&= E[f\circ X_n]E[g\circ Y_n].
\end{aligned}$$

For the converse, it suffices to consider X and Y to be simple functions and to observe that if $X = \sum_{i=1}^{l} a_i\chi_{A_i}$, where the a_i are distinct and A_i are disjoint elements of $\mathcal{F}_X$, and if $f : \mathbb{R}\to\mathbb{R}$ is the function that takes value $1/a_i$ when $x = a_i$ and 0 otherwise, then $f\circ X = \chi_{A_i}$. From this one can show that $E[AB] = E[A]E[B]$ for all $A\in\mathcal{F}_X, B\in\mathcal{F}_Y$. □

b. Conditional Probability and Conditional Expectation. Let $(\Omega,\mathcal{F},P)$ be a probability space and $Y : \Omega\to K = \{a_1, a_2, \dots\}$ a random variable taking at most countably many values. If I is any subset of K, then by definition $(Y_*P)(I) = \sum_{a_i\in I} p_i$, where $p_i := P(\{\omega\in\Omega : Y(\omega) = a_i\})$.

For each $A\in\mathcal{F}$, define a function

$$\phi(y) = \phi_A(y) = \frac{P(A\cap\{\omega : Y(\omega) = y\}}{P(\{\omega : Y(\omega) = y\})} := P(A|Y = y).$$

(This is well-defined for almost every y. More precisely, it is defined for those y for which the event $Y = y$ has nonzero probability.)

The random variable $\phi\circ Y$ is usually written $P(A|Y)$, so that $P(A|Y)(\omega) = P(A|Y = Y(\omega))$. It is called the *conditional probability of A given Y*.

Exercise 2.2.2. Let $Y = \chi_B$ be the characteristic function of a measurable set B. Then,

$$P(A|\chi_B) = P(A|B)\chi_B + P(A|B^c)\chi_{B^c}.$$

Exercise 2.2.3. Let Y be a random variable on $(\Omega,\mathcal{F},P)$ taking values in a countable set K and denote by $\mathcal{Y}$ the sub-σ-algebra of $\mathcal{F}$ generated sets of the form $Y^{-1}(L)$, where $L\subset K$. In other words, $\mathcal{Y}$ is the σ-algebra generated by Y. Show that for every $A\in\mathcal{F}$, the random variable $P(A|Y)$ is $\mathcal{Y}$-measurable and satisfies the following property: for every $E\in\mathcal{Y}$ (so $E = Y^{-1}(B)$ for some $B\subset K$),

$$\int_E P(A|Y)dP = \int_E \phi_A\circ Y dP = \int_B \phi_A d(Y_*P) = P(A\cap E).$$

Also show that if $h : \Omega\to\mathbb{R}$ is $\mathcal{Y}$-measurable and

$$\int_E h dP = P(A\cap E)$$

for all $E\in\mathcal{Y}$, then $h = P(A|Y)$ almost surely.

There is no essential reason for assuming that Y takes only countably many values other than to make the previous exercise more elementary. We discuss now the general case. Furthermore, it will be convenient and technically more flexible to define the conditional probability (and conditional expectation) in terms of a σ-algebra. (This means that we would like more generally to define $P(A|\mathcal{G})$ for a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$ in such a way that $P(A|\mathcal{G}) := P(A|Y)$ when $\mathcal{G}$ is generated by a single random variable Y.)

This is, in fact, an immediate consequence of the Radon-Nikodym theorem: if $X : \Omega \to \mathbb{R}^n$ is an integrable random variable (that is, $E[|X|] < \infty$) and $\mathcal{G}$ is any σ-algebra contained in $\mathcal{F}$, then

$$A \in \mathcal{G} \mapsto Q(A) := \int_A X dP$$

is a signed measure on $(\Omega, \mathcal{G}, P)$ which is absolutely continuous with respect to the restriction of P to $\mathcal{G}$. The Radon-Nikodym derivative of Q with respect to P is called the *conditional expectation of X given $\mathcal{G}$* and is denoted $E[X|\mathcal{G}]$. It is the unique (almost surely) $\mathcal{G}$-measurable function from Ω into $\mathbb{R}^n$ such that

$$\int_A E[X|\mathcal{G}] dP = \int_A X dP$$

for all $A \in \mathcal{G}$.

If $X = \chi_A$ is the characteristic function of a set A, we write $P(A|\mathcal{G}) := E[\chi_A|\mathcal{G}]$. If $\mathcal{G}$ is generated by random variables $X_1, \dots, X_l$, we sometimes write the conditional expectation as $E[X|X_1, \dots, X_l]$. Finally note that since $E[X|\mathcal{G}]$ is a $\mathcal{G}$-measurable random variable, the Doob-Dynkin lemma implies the existence of a measurable $\phi : \mathbb{R}^n \to \mathbb{R}$ such that $E[X|\mathcal{G}] = \phi \circ X$.

One way to think about $E[X|\mathcal{G}](\omega)$ is to regard it as the best prediction we can make about $X(\omega)$ given only the partial information about ω contained in $\mathcal{G}$. We will return later to some applications to prediction and filtering problems. Also note that it was essential for the definition of conditional expectation that X took values in a vector space. When X takes values in a manifold it becomes necessary to have some notion of barycenter.

EXERCISE 2.2.4. Show that $E[\chi_A|\chi_B] = \frac{P(A\cap B)}{P(B)}\chi_B + \frac{P(A\cap B^c)}{P(B^c)}\chi_{B^c}$

EXERCISE 2.2.5. Establish the following elementary properties of the conditional expectation. (We denote by $\langle u, v\rangle$ the dot product of $u, v \in \mathbb{R}^n$.)

(1) $E[aX + bY|\mathcal{G}] = aE[X|\mathcal{G}] + bE[Y|\mathcal{G}]$
(2) $E[E[X|\mathcal{G}]|\mathcal{H}] = E[X|\mathcal{H}]$ if $\mathcal{H} \subset \mathcal{G}$
(3) $E[X|\mathcal{G}] = X$ if X is $\mathcal{G}$-measurable
(4) $E[X|\mathcal{G}] = E[X]$ if X is independent of $\mathcal{G}$

EXERCISE 2.2.6. Show that the conditional expectation behaves like a (function valued) probability measure. More precisely, it satisfies:

(1) $P(A|\mathcal{G}) \geq 0$
(2) $P(\Omega|\mathcal{G}) = 1$
(3) If $A_1, A_2, \dots$ are mutually disjoint, then $P(\cup_{n=1}^{\infty} A_n|\mathcal{G}) = \sum_{n=1}^{\infty} P(A_n|\mathcal{G})$.

EXERCISE 2.2.7. Suppose that $X \in L^2(\Omega, \mathcal{F}, P)$ and that $\mathcal{G} \subset \mathcal{F}$ is a sub-σ-algebra. Show that $E[X|\mathcal{G}]$ is the orthogonal projection of X on the closed subspace $L^2(\Omega, \mathcal{G}, P)$.

The previous exercise shows that the conditional expectation solves the problem of finding an approximation of X using the information encoded in $\mathcal{G}$ such that the mean square error is minimized. In other words, the conditinal expectation is the $\mathcal{G}$-measurable random variable Y that minimizes $E[(X-Y)^2]$. It can thus be regarded as the best *prediction* we can make of X given the information in $\mathcal{G}$.

c. Conditional Independence. The notion of statistical independence can be generalized as follows. Two random variables X, Y are said to be *conditionally independent* with respect to a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$ if for every B_1 in the σ-algebra $\mathcal{X}$ generated by X and every B_2 in the σ-algebra $\mathcal{Y}$ generated by Y we have

$$P[B_1 \cap B_2|\mathcal{G}] = P[B_1|\mathcal{G}]P[B_2|\mathcal{G}].$$

Equivalently, for every $\mathcal{X}$-measurable Z_1 and every $\mathcal{Y}$-measurable Z_2,

$$E[Z_1 Z_2|\mathcal{G}] = E[Z_1|\mathcal{G}]E[Z_2|\mathcal{G}].$$

d. Jensen's Inequality. A function $\phi : \mathbb{R} \to \mathbb{R}$ is said to be *convex* if for all $x, y \in \mathbb{R}$ and $\lambda \in [0,1]$,

$$\phi(\lambda x + (1-\lambda)y) \leq \lambda\phi(x) + (1-\lambda)\phi(y).$$

The following is known as *Jensen's inequality*: If ϕ is convex, then for every X and $\mathcal{G}$,

$$E[\phi(X)|\mathcal{G}] \geq \phi(E[X|\mathcal{G}]).$$

e. Conditional Expectation for Random Variables with Densities. In this section we derive expressions for the conditional expectations when the random variables have probability densities.

Thus suppose that $X : \Omega \to \mathbb{R}^n$ and $Y : \Omega \to \mathbb{R}^m$ are random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ and that the law of $X \times Y$ has a density $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ (with respect to the Lebesgue measure). This means that f is a nonnegative function having total integral 1 and

$$d[(X \times Y)_* P] = f(x,y)dxdy.$$

The probability densities of X and Y are, by definition, the functions $f_X : \mathbb{R}^n \to \mathbb{R}$ and $f_Y : \mathbb{R}^m \to \mathbb{R}$ such that

$$f_X(x)dx = d(X_* P), \;\; f_Y(y)dy = d(Y_* P).$$

Exercise 2.2.8. Show that the following identities hold:

$$f_X(x) = \int_{\mathbb{R}^m} f(x,y)dy$$
$$f_Y(y) = \int_{\mathbb{R}^n} f(x,y)dx.$$

(Note: if $\pi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is the natural projection, then $X_* P = \pi_*(X \times Y)_* P$. Use this to show that $\int_A f_X(x)dx = \int_A \int_{\mathbb{R}^m} f(x,y)dydx$.)

For simplicity we will assume that $m = n = 1$, although it will be clear that the formulas that will result are general. The Borel σ-algebra of $\mathbb{R}$ will be denoted $\mathcal{B}$.

EXERCISE 2.2.9. Let $B_1, B_2 \in \mathcal{B}$, set $A = X^{-1}(B_1)$ and let $\phi(y) = P(A|Y = y)$. Show that

$$\int_{B_1} \phi(y) f_Y(y) dy = \int_{B_2} \int_{B_1} f(x, y) dx dy.$$

Given the previous exercise, we conclude that

$$P(X \in B|Y = y) = \int_B f(x|y) dx,$$

where

$$f(x|y) := \frac{f(x, y)}{f_Y(y)}.$$

We call $x \mapsto f(x|y)$ the probability density of X given $Y = y$.

For a concrete interpretation of the function $f(x|y)$, think of the following situation. Suppose that the random variable X gives the height of people taken from a certain population, and Y gives the weight. Suppose that the joint probability distribution is $f(x, y)$. Then $x \mapsto f(x|y)$ is the probability density of heights of people whose weight is y.

EXERCISE 2.2.10 (Bayes' formula 1). If $B_1, B_2, \ldots$ is a countable or finite measurable partition of Ω such that $P(B_i) \neq 0$ for all i, show that

$$P(B_n|A) = \frac{P(A|B_n)P(B_n)}{\sum_{i=1}^{\infty} P(A|B_i)P(B_i)}.$$

EXERCISE 2.2.11 (Bayes' formula 2). Show that

$$f(y|x) = \frac{f_Y(y)}{f_X(x)} f(x|y).$$

EXERCISE 2.2.12 (Bayes' formula 3). Assume that P and Q are two probability measures on $(\Omega, \mathcal{F})$ and that $dP = LdQ$. Denote by $E^P[X|\mathcal{G}]$ (respectively, $E^Q[X|\mathcal{G}]$) the conditional expectation with respect to the measure P (respectively, Q). Show that

$$E^P[X|\mathcal{G}] = \frac{E^Q[XL|\mathcal{G}]}{E^Q[L|\mathcal{G}]}.$$

f. Conditioning and Disintegration.

3. Examples of probability spaces

a. Uniform Distribution on the Unit Interval. The unit interval $[0, 1]$ with the Lebesgue measure P_0 and the completion of the Borel σ-algebra is one of the simplest and most fundamental probability spaces. All the other spaces studied in these notes are obtained by considering different random variables on $[0, 1]$. One practical consequence of this remark is that knowning how to numerically simulate the uniform distribution on the unit interval, one can, in principle, simulate any distribution on other (Polish) spaces.

For example, consider a finite probability space, $\Omega = \{1, \ldots, n\}$. The family of all subsets of Ω will be taken as the σ-algebra and P is specified by the values $P(\{i\}) = p_i > 0$, $p_1 + \cdots + p_n = 1$. Choose a partition of $[0, 1]$ into intervals I_i of length p_i and define a random variable $X : [0, 1] \to \Omega$ as the simple function that takes value i over the interval I_i. Then $P = X_* P_0$.

As a second example of how other probability spaces arise as images of appropriate random variables on the unit interval, let $\Omega = \mathbb{R}$ and suppose that P is given by a smooth positive density $\rho : \mathbb{R} \to (0, \infty)$. This means that the probability of $A \subset \mathbb{R}$ is given by

$$P(A) = \int_A \rho(t)\ dt.$$

Consider the smooth function $Z : \mathbb{R} \to (0,1)$ given by

$$Z(x) := \int_{-\infty}^{x} \rho(t)\ dt.$$

Since $Z'(x) = \rho(x) > 0$ for every x, by the inverse function theorem Z is invertible. Let $X : (0,1) \to \mathbb{R}$ denote its inverse. Then $X_* P_0 = P$. In fact, for each $x \in \mathbb{R}$,

$$\begin{aligned} P((-\infty, x)) &= \int_{-\infty}^{x} \rho(t)\ dt \\ &= Z(x) \\ &= P_0((0, Z(x))) \\ &= P_0(X^{-1}(-\infty, x)) \\ &= X_* P_0((-\infty, x)). \end{aligned}$$

Thus, $P = X_* P_0$ on sets of the form $(-\infty, x)$. But these sets, together with taking countable unions and complements, generate the Borel σ-algebra of $\mathbb{R}$. Therefore, $P = X_* P_0$.

Other examples will be studied in the subsequent sections.

b. Coin tossing. Let $\Omega = \{0,1\}^{\mathbb{N}} = \{0,1\} \times \{0,1\} \times \dots$ be the set of all sequences $\omega = (a_1, a_2, \dots)$, with i-th coordinate $a_i \in \{0,1\}$. We denote the i-th coordinate function by $x_i : \Omega \to \{0,1\}$, $x_i(\omega) = a_i$. An element of Ω can be regarded as an outcome of the experiment of tossing a coin an infinite number of times.

Let $\mathcal{F}_n$ denote the σ-algebra of all subsets of Ω of the form $A \times \{0,1\}^{\mathbb{N}}$, where $A \subset \{0,1\}^n$. In other words, elements of $\mathcal{F}_n$ are subsets of Ω for which only the coordinates up to x_n are restricted. Notice that $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots$. Define $\mathcal{F}$ as the smallest σ-algebra of Ω that contains $\mathcal{F}_n$ for all n. Also, for a fixed $p \in [0,1]$, define a probability measure $P : \mathcal{F}_n \to [0,1]$ so that, writing $p_0 = p, p_1 = 1 - p$,

$$P(\{\omega \in \Omega : x_1(\omega) = a_1, \dots, x_n(\omega) = a_n\}) = p_{a_1} \dots p_{a_n}.$$

Exercise 2.3.1. Using Kolmogorov's extension theorem, show that P extends uniquely to a probability measure on Ω and that $(\Omega, \mathcal{F}, P)$ defines a probability space.

Exercise 2.3.2. Show that the projections $x_i : \Omega = \{0,1\}^{\mathbb{N}} \to \{0,1\}$ are all $\mathcal{F}$-measurable. For the same example, show that a function $f : \Omega \to \mathbb{R}$ is $\mathcal{F}_n$-measurable exactly when there is a function $g : \{0,1\}^n \to \mathbb{R}$ such that $f(\omega) = g(x_1(\omega), \dots, x_n(\omega))$. (Doob-Dynkin lemma.)

Exercise 2.3.3. Let $(I, \mathcal{B}, \lambda)$ denote the probability space where $I = [0,1]$, $\mathcal{B}$ is the Borel σ-algebra (which is generated by intervals), and λ is the Lebesgue measure on I. Let $(\Omega, \mathcal{F}, P)$ be the (fair) coin-tossing probability space (so that P is defined for $p = 1/2$). Define the map $f : \Omega \to I$ defined by

$$f(a_1, a_2, \dots) = \frac{a_1}{2} + \frac{a_2}{2^2} + \dots + \frac{a_n}{2^n} + \dots.$$

(1) Show that f is measurable. (It suffices to show that $f^{-1}(I) \in \mathcal{F}$ for all dyadic intervals $I = [k/2^n, (k+1)/2^n]$.)
(2) Show that the pull-back of P under f is λ. (Once again, it suffices to show that $f_*P(I) = 1/2^n$ for dyadic intervals of the form $I = [k/2^n, (k+1)/2^n]$.)
(3) Let $Q = \{k/2^n : n \in \mathbb{N}, 0 \leq k \leq 2^n\}$ and $E = f^{-1}(Q)$. Show that
   (a) E is the set of $\omega \in \Omega$ such that $f(\omega)$ has more than one pre-image.
   (b) Show that E is the set of $\omega = (a_1, a_2, \dots) \in \Omega$ such that, for some N sufficiently large, $a_N = a_{N+1} = a_{N+2} = \dots$.
   (c) Show that E and Q are measurable, $P(E) = 0 = \lambda(Q)$, and that
$$f : \Omega - E \to I - Q$$
   is a bijection.

The previous exercise has the following corollary.

Proposition 2.3.4. The coin-tossing probability space is isomorphic to $(I, \mathcal{B}, \lambda)$.

The above proposition suggests a mathematical model for the tossing-a-fair-coin experiment: pick a number in $[0, 1]$ at random (for the uniform distribution) and look at the number's dyadic expansion coefficients. The n-th coefficient is the outcome of the n-th toss of the coin.

Exercise 2.3.5. Show that $\{0, 1\}^{\mathbb{N}}$, with the product topology, is homeomorphic with the middle-third Cantor set.

We recall for the sake of the previous exercise that any compact, totally disconnected topological space is homeomorphic to the middle-third Cantor set. The latter can be described as the set of all $x \in [0, 1]$ with expansion
$$x = \frac{a_1}{3} + \frac{a_2}{3^2} + \frac{a_3}{3^3} + \dots$$
such that a_i is either 0 or 2 for all i.

c. A Little Ergodic Theory Inspired by Coin-tossing. In the coin-tossing experiment, one could also consider $\Omega = \{0, 1\}^{\mathbb{Z}}$, the set of all doubly infinite sequences of 0 and 1. In this case, the experiment runs from the infinite past into the infinite future. We assume now that this is the case. The σ-algebra is more conveniently described as the one generated by the *cylinder sets*, that is, by subsets of Ω of the form
$$C = \{\omega \in \Omega \mid \omega(i) = a_i, \ r \leq i \leq s\}$$
where a_i are in $\{0, 1\}$.

Let p_0, p_1, be positive numbers such that $p_0 + p_1 = 1$. Then, there exists a unique measure P on $(\Omega, \mathcal{F})$ such that on a cylinder set such as C above,
$$P(C) = p_{a_r} p_{a_{r+1}} \cdots p_{a_s}.$$

Define the *shift map* $S : \Omega \to \Omega$ by
$$(S\omega)(i) = \omega(i + 1), \quad i \in \mathbb{Z}.$$
S is a (bijective) isomorphism of Ω and with respect to it the outcome of the coin-tossing experiment at time k is given by the random variable $\pi_0 \circ S^k$, in which $\pi_0 : \Omega \to \{0, 1\}$ represents the 0th coordinate function on Ω.

In a previous exercise you were asked to show that the coin-tossing process (with $\Omega = \{0, 1\}^{\mathbb{N}}$) is (up to sets of measure 0) isomorphic to the Lebesgue probability space consisting of the unit interval $[0,1]$ with the Lebesgue measure. In fact, an isomorphism can be defined by

$$\Phi : (\omega_1, \omega_2, \dots) \mapsto 0.\omega_1\omega_2 \dots .$$

Under this isomorphism, the shift map S corresponds to the transformation of the interval

$$S(x) = \begin{cases} 2x & \text{if } 0 \leq x < 1/2 \\ 2x - 1 & \text{if } 1/2 \leq x \leq 1. \end{cases}$$

(Equivalently, we could consider the map $Q : z \mapsto z^2$ from the unit circle in $\mathbb{C}$ to itself. The correspondence between S and Q is clear if we regard the map $x \mapsto e^{2\pi i x}$ from the interval to the circle.)

If $\Omega = \{0,1\}^{\mathbb{Z}}$, the shift map is invertible. One obtains an isomorphism of Ω with the unit square $[0,1] \times [0,1]$ by the assignment

$$(\cdots \omega_{-1}\omega_0\omega_1 \cdots) \mapsto (0.\omega_0\omega_1\omega_2 \cdots, 0.\omega_{-1}\omega_{-2} \cdots).$$

Under this isomorphism, the shift map corresponds to the *baker's transformation* of the unit square:

$$B : (x, y) \mapsto (S(x), \frac{1}{2}(y + \chi_{I_1}(x)))$$

where χ_{I_1} is the characteristic function of the interval $[1/2, 1]$

Exercise 2.3.6. Explain why the transformation is called the *baker's transformation* by giving a geometric description of what it does to the unit square.

Anosov Diffeo of 2-torus and isomorphism with Markov chain

For the next exercise, we make the following definition. Let T be a measure preserving transformation of a probability space $(\Omega, \mathcal{F}, P)$. Then T is said to be *ergodic* if for any invariant set $A \in \mathcal{F}$ (that is, such that $T^{-1}(A) = A$) either $P(A) = 0$ or $P(A) = 1$. This condition is equivalent to the following: if f is a T-invariant function, that is, $f \circ T = f$ almost surely, then f is constant almost surely.

Exercise 2.3.7. Show that the shift-map is ergodic. (Suggestion: show first that the map $z \mapsto z^2$ of the circle is ergodic by proving that the Fourier series of the characteristic function of an invariant set only contains the constant term. Then use the fact the the $z \mapsto z^2$ is isomorphic to the semi-infinite shift. A similar argument works for the baker's transformation and the doubly-infinite shift.)

For each $\omega \in \Omega = \{0,1\}^{\mathbb{N}}$, denote by $S_n(\omega)$ the number of times that 1 appears in the finite sequence $x_1(\omega), x_2(\omega), \dots, x_n(\omega)$. Of course, the sequence $S_n(\omega)/n$ need not have any limit as $n \to \infty$. However, the *strong law of large numbers* asserts that the limit exists and is equal to $1/2$ almost surely. Another way in which this property can be expressed is as follows. Let $f : \Omega \to \mathbb{R}$ be the first coordinate map, $f(\omega) = x_1(\omega)$ and let S, as before, denote the shift map. Then for each ω,

$$\frac{S_n(\omega)}{n} = \frac{1}{n} \sum_{i=0}^{n-1} f(S^i(\omega)).$$

This leads to the investigation of limits $(1/n)\sum_{i=0}^{n-1} f \circ S^i$, where S is a measure preserving transformation and f is a measurable function on some probability space Ω. The main result about the existence of such limits is *Birkhoff's ergodic theorem.*

The next theorem is a fundamental result in ergodic theory.

THEOREM 2.3.8 (Birkhoff's ergodic theorem). Let S be a measure preserving transformation of a probability space $(\Omega, \mathcal{F}, P)$ and let f be an integrable function on Ω. Then the sequence $(1/n)\sum_{i=0}^{n-1} f(S^i(\omega))$ converges for almost every ω to $\bar{f}(\omega)$, for some integrable S-invariant function $\bar{f}$. Furthermore,

$$\int_\Omega \bar{f} dP = \int_\Omega f dP.$$

If S is ergodic, then Birkhoff's theorem implies that $(1/n)\sum_{i=0}^{n-1} f \circ S^i$ converges almost surely to the integral of f over Ω.

It should now be clear how Birkhoff's theorem implies the law of large numbers for the coin-tossing experiment. Using $\Omega = [0,1]$, P the Lebesgue measure, and S the map that takes the fractional part after multiplication by 2, to describe the coin-tossing experiment, then the n-th toss gives 1 precisely when $S^n(\omega)$ lies in the half-interval $[1/2, 1]$, so the frequency of 1 in a trial ω is given by the limit of $1/n \sum_{i=0}^{n-1} \chi_{[1/2,1]}(S^i(\omega))$. Since the transformation is ergodic, the limit is the integral $\int_0^1 \chi_{[1/2,1]} dP = 1/2$, almost surely.

d. Getting Countably Many Independent Random Variables. It is possible to obtain a random variable from $[0,1]$ to $[0,1]^{\mathbb{N}}$ that sends the Lebesgue measure to the product Lebesgue measure on the infinite cube, in such a way that its coordinate projections onto $[0,1]$ are independent random variables. We show in this section how this can be done.

Let $a := (a_1, a_2, \dots)$ be a sequence of positive integers and define a function $\varphi_a : [0,1] \to [0,1]$ by

$$0.x_1 x_2 \cdots \mapsto 0.x_{a_1} x_{a_2} \cdots$$

where $0.x_1 x_2 \dots$ is the unique binary expansion of x with infinitely many zeros. This is a measurable function. Viewed as a function from $\Omega = \{0,1\}^{\mathbb{N}}$ into itself, φ_a is simply

$$\omega \mapsto (x_{a_1}(\omega), x_{a_2}(\omega), \dots)$$

where the $x_i : \Omega \to \{0,1\}$ are the (measurable) projections.

PROPOSITION 2.3.9. Let P denote the Lebesgue measure on $[0,1]$. Then $(\varphi_a)_* P = P$ for any sequence a. In other words, the random variable φ_a is uniformly distributed. Furthermore, if $b \subset \mathbb{N}$ is another infinite sequence such that a and b are disjoint, then φ_a and φ_b are independent random variables.

PROOF. The fact that φ_a and φ_b are independent is an immediate consequence of the fact that the projections x_i are independent. To prove that φ_a is uniformly distributed, it suffices to show that

$$P(\{x \in [0,1] \mid \varphi_a(x) < \alpha\}) = \alpha$$

for each $\alpha \in [0,1]$. Writing $\alpha = 0.\alpha_1 \alpha_2 \cdots$, we have that $\varphi_a^{-1}([0,\alpha])$ decomposes as a disjoint union of the measurable sets:

$$\begin{aligned} &I_1 := \{x \in [0,1] \mid x_{a_1} < \alpha_1\} \\ &I_2 := \{x \in [0,1] \mid x_{a_1} = \alpha_1 \text{ and } x_{a_2} < \alpha_2\} \\ &I_3 := \{x \in [0,1] \mid x_{a_1} = \alpha_1 \text{ and } x_{a_2} = \alpha_2 \text{ and } x_{a_3} < \alpha_3\} \\ &\quad \dots \end{aligned}$$

Note that

$$P(\{x \in [0,1] \mid x_{a_i} < \alpha_i\}) = \frac{\alpha_i}{2}$$
$$P(\{x \in [0,1] \mid x_{a_i} = \alpha_i\}) = 1/2.$$

This, together with the independence of the x_i, imply:

$$P(I_1) = \alpha_1/2$$
$$P(I_2) = \alpha_2/2^2$$
$$P(I_3) = \alpha_3/2^3$$
$$\cdots$$

Therefore,

$$(\varphi_a)_* P([0,\alpha]) = \frac{\alpha_1}{2} + \frac{\alpha_2}{2^2} + \frac{\alpha_3}{2^3} + \cdots = \alpha.$$

□

e. Countably Many Independent Gaussian Random Variables. The standard Gaussian (or normal) distribution on $\mathbb{R}$ (with mean 0 and variance 1) is the probability measure

$$P(A) = (2\pi)^{-1/2} \int_A e^{-\frac{x^2}{2}}\, dx.$$

On $\mathbb{R}^n$, the standard normal distribution is the product measure:

$$P(A) = (2\pi)^{-n/2} \int_A e^{-\frac{|x|^2}{2}}\, dx.$$

Like we did in the previous section for the uniform distribution, we can construct a countable sequence of independent identically distributed normal random variables as follows. In fact, let $X : (0,1) \to \mathbb{R}$ be a random variable whose probability distribution has density $\rho(x) = (2\pi)^{-n/2} e^{-\frac{|x|^2}{2}}$ and let

$$\Phi = (\Phi_1, \Phi_2, \dots) : (0,1) \to (0,1)^{\mathbb{N}}$$

be the map constructed in the previous section. Then $G = (X \circ \Phi_1, X \circ \Phi_2, \dots)$ is a sequence of independent normal random variables.

f. Random Variables and Affine Transformations. This section describes a few general properties of Gaussian normal variables. The Lebesgue measure on $\mathbb{R}^n$ will be denoted by λ. For simplicity we will write $d\lambda(x) = dx$. We say that a map $T : \mathbb{R}^n \to \mathbb{R}^n$ is an affine transformation if $T(x) = Ax + b$, where A is a linear map and b is a vector. Here x denotes a point in $\mathbb{R}^n$, viewed as a column vector. We will regard A as an $n \times n$ matrix.

Lemma 2.3.10. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an invertible affine transformation of the form $T(x) = Ax + b$. Then for any integrable function $f : \mathbb{R}^n \to \mathbb{R}$, the Radon-Nikodym derivative of $T_*\lambda$ with respect to λ is*

$$\frac{d(T_*\lambda)}{d\lambda} = |\det A|^{-1}.$$

PROOF. We write $T = A \circ \tau$, where $\tau : x \mapsto x + b$. Now by the Gram-Schmidt orthogonalization procedure an invertible matrix A can be written as a product $A = BC$, where B is upper-triangular and C is orthogonal. Furthermore, a simple induction argument shows that an upper-triangular matrix B decomposes as the product DU of a diagonal matrix D and an upper-triangular matrix U with diagonal entries all equal to 1. It is then possible to write $T = D \circ U \circ C \circ \tau$. But translations, linear isometries, and shears (as the linear transformation defined by U) do not change the Lebesgue measure, so $T_* \lambda = D_* \lambda$. On the other hand, it is elementary to show that $D_* \lambda = |\det D|^{-1} \lambda = |\det A|^{-1} \lambda$. Consequently, $T_* \lambda = |\det A|^{-1} \lambda$, and $|\det A|^{-1}$ must be the Radon-Nikodym derivative as claimed. □

The family of all invertible affine transformations of $\mathbb{R}^n$ forms a group, which we will denote by $\mathrm{Aff}(n, \mathbb{R})$.

EXERCISE 2.3.11. Write elements of $\mathrm{Aff}(n, \mathbb{R})$ in the form (A, b). Show:

(1) The identity element is $(I, 0)$, where I is the identity $n \times n$ matrix.
(2) $(A_1, b_1)(A_2, b_2) = (A_1 A_2, A_1 b_2 + b_1)$
(3) $(A, b)^{-1} = (A^{-1}, -A^{-1} b)$
(4) The subgroup consisting of pure translations (that is, of elements (I, b)) is a normal subgroup. (In fact, $\mathrm{Aff}(n, \mathbb{R})$ is a semidirect product of the general linear group and the group of translations.)
(5) The map
$$T \in \mathrm{Aff}(n, \mathbb{R}) \mapsto \frac{d(T_* \lambda)}{d\lambda} \in (0, \infty)$$
is a group homomorphism from $\mathrm{Aff}(n, \mathbb{R})$ into the multiplicative group of positive reals.

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $A = (a_{ij})$ be a positive symmetric $n \times n$ matrix and define the quadratic form $Q(x) = \frac{1}{2} x^t A x$, where x^t denotes transpose.

A random variable $X : \Omega \to \mathbb{R}^n$ is called a *normal, or gaussian* random variable if its probability distribution has the form $\rho(x) = \text{constant} e^{-Q(x - x_0)}$ for some vector $x_0 \in \mathbb{R}^n$ and some quadratic form Q. The constant is determined by normalization.

EXERCISE 2.3.12. Show that the normalization constant is $(2\pi)^{-n/2} (\det A)^{1/2}$. (As in a previous argument it can be assumed that A is diagonal.)

EXERCISE 2.3.13. If $X = (X_1, \dots, X_n)$ is a gaussian random variable with parameters (x_0, V), where $V = A^{-1}$, show that $E[X] = x_0$ and that V is the covariance matrix of $X_1, \dots, X_n$.

Representing ρ by the pair (x_0, V), $V = A^{-1}$, the collection of all gaussian probability densities may be identified with the set (manifold) $M := \mathbb{R}^n \times S$, where S is the set of all positive definite symmetric $n \times n$ matrices. We think of S itself as the set of all *centered gaussian distributions*. (S is a homogeneous space of the form $SL(n, \mathbb{R})/SO(n)$.)

The group $\mathrm{Aff}(n, \mathbb{R})$ is a group of transformations of M by defining the action
$$(A, b)(x_0, V) := (A x_0 + b, A V A^t).$$

EXERCISE 2.3.14. Show that this indeed defines a group action. That is, show that the map
$$\Phi : \mathrm{Aff}(n, \mathbb{R}) \times M \to M$$
defined by $\Phi((A, b), (x_0, V)) := (A x_0 + b, A V A^t)$ satisfies the properties:

(1) $\Phi(e,\xi) = \xi$, where e is the identity element in $\mathrm{Aff}(n,\mathbb{R})$ and $\xi \in M$;
(2) For all $g_1, g_2 \in \mathrm{Aff}(n,\mathbb{R})$, and $\xi \in M$,

$$\Phi(g_1 g_2, \xi) = \Phi(g_1, \Phi(g_2, \xi)).$$

extend next proposition to case $T : \mathbb{R}^n \to \mathbb{R}^m$ for $m \leq n$.

Proposition 2.3.15. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, and denote by $N(x_0, V)$ the set of random variables $X : \Omega \to \mathbb{R}^n$ with mean x_0 and covariance matrix V. Let $T(x) = Cx + b$ be an invertible affine transformation of $\mathbb{R}^n$. Then for every $X \in N(x_0, V)$, we have $T \circ X \in N(Cx_0 + b, CVC^t)$.*

Proof. We write $d(X_*P)(x) = \rho(x)dx$. Recall that λ is our notation for the Lebesgue measure and that $d\lambda(x) = dx$. For any integrable function f we have

$$\begin{aligned}\int_{\mathbb{R}^n} f(x)d(T_*X_*P)(x) &= \int_{\mathbb{R}^n} f(T(x))d(X_*P)(x)\\ &= \int_{\mathbb{R}^n} f(T(x))\rho(x)d\lambda(x)\\ &= \int_{\mathbb{R}^n} f(x)\rho(T^{-1}(x))d(T_*\lambda)(x)\\ &= \int_{\mathbb{R}^n} f(x)\rho(T^{-1}(x))|\det C|^{-1}dx.\end{aligned}$$

Since f is arbitrary we conclude that T_*X_*P has density

$$h(x) = \rho(T^{-1}(x))|\det C|^{-1}.$$

To conclude, use $|\det V|^{1/2}|\det C| = |\det(CVC^t)|$ and $T^{-1}(x) = C^{-1}(x - b)$. □

Proposition 2.3.16. *If $X : \Omega \to \mathbb{R}^n$ and $Y : \Omega \to \mathbb{R}^m$ are gaussian random variables, then they are independent if and only if they are uncorrelated (that is, $cov(X_i, Y_j) = 0$ for all i, j.*

Proof. Independent random variables are always uncorrelated. Conversely, if X and Y are uncorrelated gaussian random variables, then because of the form of the covariance matrix of (X, Y), the probability density of (X, Y) factors as a product $\rho_1(x)\rho_2(y)$. But this is equivalent to independence. □

g. Gaussian Families. See page 2 and 3 of Stroock's book "An Introduction to the Analysis of Paths on a Riemannian Manifold."

4. Random Walks and the Central Limit Theorem

a. Random Walk on the Real Line. Let $(\Omega, \mathcal{F}, P)$ denote the coin-tossing probability space. (In particular, $\Omega = \{0,1\}^{\mathbb{N}}$.) Rather than work with the coordinates x_i as before, it will be more convenient here to use $\pi_i : \Omega \to \{-1, 1\}$, such that $\pi_i(\omega) = 1$ if $x_i(\omega) = 0$ and $\pi_i(\omega) = -1$ if $x_i(\omega) = 1$.

Fix $n \in \mathbb{N}$ and define a motion on the real line with velocity given by the following random function of t:

$$v^{(n)}(t) := \sqrt{n}\pi_{[nt]+1} : \Omega \to \{-\sqrt{n}, \sqrt{n}\}.$$

In other words, $v^{(n)}(t) = \sqrt{n}\pi_k$ when t is in the interval $[(k-1)/n, k/n)$. The motion on $\mathbb{R}$ is obtained by integrating the velocity process. It is described by the random process $x^{(n)}(t)$, $t \geq 0$, such that $x^{(0)}$ will be chosen to be 0 and

$$x^{(n)}(t) = \frac{1}{\sqrt{n}}\left\{\pi_1 + \pi_2 + \cdots + \pi_{[nt]} + (nt - [nt])\pi_{[nt]+1}\right\}.$$

Note that

$$x^{(n)}(\frac{m}{n}) = \frac{1}{\sqrt{n}}\{\pi_1 + \pi_2 + \cdots + \pi_m\} = \sqrt{\frac{m}{n}}\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\pi_i.$$

We now fix t and consider what happens to $x^{(n)}(t)$ as $n \to \infty$ and $\frac{m}{n} \to t$. The *central limit theorem* immediately implies that the probability distribution of $x^{(n)}(t)$ converges to a Gaussian probability distribution. Thus we have:

PROPOSITION 2.4.1. *As $n \to \infty$,*

$$P(a \leq x^{(n)}(t) \leq b) \to \int_{a/\sqrt{t}}^{b/\sqrt{t}} \frac{e^{\frac{x^2}{2}}}{\sqrt{2\pi}}dx = \int_a^b \frac{e^{\frac{x^2}{2t}}}{\sqrt{2\pi t}}dx.$$

PROOF. This is a consequence of the central limit theorem, which states that

$$P(a \leq \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\pi_i \leq b) \to \int_a^b \frac{e^{\frac{x^2}{2}}}{\sqrt{2\pi}}dx$$

as $n \to \infty$. □

b. Khinchine's Proof of the Central Limit Theorem. We sketch in this section a simple proof of the central limit theorem due to Khinchine, from 1933.

Consider the function

$$u(x,t) = \frac{1}{\sqrt{2\pi t}}\int_{-\infty}^{+\infty} e^{-\frac{(x-y)^2}{2t}} f(y)dy.$$

It is a simple exercise to verify that u is a solution of the heat equation:

$$u_t = \frac{1}{2}u_{xx},$$

for $t > 0$ and $x \in \mathbb{R}$, and satisfying the initial condition $u(x,0) = f(x)$.

Define

$$x^{(n)}(\frac{m}{n}) = \frac{1}{\sqrt{n}}\{\pi_1 + \pi_2 + \cdots + \pi_m\}.$$

Also define, for $t = m/n$, the functions

$$u^{(n)}\left(\frac{k}{\sqrt{n}}, t\right) = E[f(x^{(n)}(t))|x^{(n)}(0) = \frac{k}{\sqrt{n}}].$$

Then $u^{(n)}$ satisfies the equation

$$u^{(n)}\left(\frac{k}{\sqrt{n}}, \frac{m}{n}\right) = \sum_{j=-\infty}^{+\infty} p(k,j;m)f(\frac{j}{\sqrt{n}}),$$

where $p(k,j;m)$ is the probability that, having started at $k/\sqrt{n}$ at time 0, the particle describing a random walk along the line will find itself at $j/\sqrt{n}$ at time m/n. Thus

$$p(k,j;m) = P(\pi_1 + \cdots + \pi_m = j - k).$$

EXERCISE 2.4.2. Show that

$$p(k,j;m+1) = \frac{1}{2}[p(k+1,j;m) + p(k-1,j;m)].$$

A different way to express the equation in the previous exercise is to write

$$p(k,j;m+1) - p(k,j;m) = \frac{1}{2}[p(k+1,j;m) - 2p(k,j;m) + p(k-1,j;m)].$$

Notice how this looks like a discretized form of the heat equation. In fact, the proof will proceed by comparing solutions of the continuous and discrete equations. We first set some notation to facilitate this comparison. Write $\delta = 1/\sqrt{n}$, $\tau = 1/n$, and define $u_{xx}^{(n)} = \Delta^{(n)} u^{(n)}$, where

$$\Delta^{(n)} h\,(k\delta, m\tau) = \frac{h((k+1)\delta, m\tau) - 2h(k\delta, m\tau) + h((k-1)\delta, m\tau)}{\delta^2},$$

which is a discretized second derivative in x. The first derivative in t has the following discrete form:

$$u_t^{(n)}(k\delta, m\tau) = \frac{u^{(n)}(k\delta,(m+1)\tau) - u^{(n)}(k\delta, m\tau)}{\tau}.$$

We now have:

$$\begin{aligned}
\tau u_t^{(n)}(k\delta, m\tau) &= \sum_{j=-\infty}^{+\infty} [p(k,j;m+1) - p(k,j;m)] f(j\delta) \\
&= \sum_{j=-\infty}^{+\infty} \frac{1}{2}[p(k+1,j;m) - 2p(k,j;m) + p(k-1,j;m)] f(j\delta) \\
&= \frac{1}{2}\delta^2 u_{xx}^{(n)}(k\delta, m\tau).
\end{aligned}$$

Since $\delta^2 = \tau$, we have

$$u_t^{(n)}(k\delta, m\tau) = \frac{1}{2} u_{xx}^{(n)}(k\delta, m\tau).$$

Our goal is to show that

$$u^{(n)}(\frac{k}{\sqrt{n}}, \frac{m}{n}) \to u(x,t)$$

as $n \to \infty$, $m/n \to t$, and $k/\sqrt{n} \to x$.

Since u satisfies the heat equation, then

$$\frac{u(k\delta, \frac{m}{n} + \tau) - u(k\delta, \frac{m}{n})}{\tau} = \Delta^{(n)} u(k\delta, \frac{m}{n}) + o(1).$$

Here, $o(1) \to 0$ uniformly in k and m as $n \to \infty$.

After some algebraic simplification, we write the discretized equations in a recursive form:

$$\begin{aligned}
u(k\delta, \frac{m}{n} + \tau) &= \frac{1}{2}\left[u((k+1)\delta, \frac{m}{n}) + u((k-1)\delta, \frac{m}{n})\right] + \tau o(1) \\
u^{(n)}(k\delta, \frac{m}{n} + \tau) &= \frac{1}{2}\left[u^{(n)}((k+1)\delta, \frac{m}{n}) + u^{(n)}((k-1)\delta, \frac{m}{n})\right].
\end{aligned}$$

Using the initial condition $u^{(n)}(k\delta, 0) = f(k\delta) = u(k\delta, 0)$ and the above recursive formulas we finally get

$$u^{(n)}(k\delta, \frac{m}{n}) - u(k\delta, \frac{m}{n}) = \frac{m}{n} o(1).$$

We can now pass to the limit to conclude the proof.

c. A Geometric Interpretation of the Central Limit Theorem. Let $\mathcal{C}$ be the subset of $\mathbb{R}^n$ that consists of all the vertex points of the n-dimensional unit cube $[0,1]^n$. Note that $\mathcal{C}$ has 2^n elements. Let V be the vertex that is farthest away from the origin (which is also a vertex), and C the center of the cube. The center of the cube corresponds to the vector $c = \frac{1}{2}\overrightarrow{OV}$. Define the unit vector $u = \frac{\overrightarrow{OV}}{|\overrightarrow{OV}|}$. Also define the map $\pi : \mathbb{R}^n \to \mathbb{R}$ such that $\pi(x) = (x - c) \cdot u$ (the ordinary dot product of the two vectors). This is the orthogonal projection of the translate $x - c$ along the direction from O to V.

Let μ_n be the measure on $\mathcal{C}$ that assigns mass $1/2^n$ to each vertex and let $P_n := \pi_* \mu_n$. Then, as $n \to \infty$, P_n converges to the standard Gaussian measure on the real line.

d. More General Random Walks; Random Flights.

CHAPTER 3

Differential Calculus in $\mathbb{R}^n$

1. Differentiable Maps on $\mathbb{R}^n$

Elements of $\mathbb{R}^n$ will often be written as column vectors. The standard basis of $\mathbb{R}^n$ is thus

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Given a vector $v = \sum_{i=1}^n a_n e_i \in \mathbb{R}^n$, its (Euclidean) *norm* will be written

$$|v| = (v_1^2 + \cdots + v_n^2)^{1/2}.$$

Let $M(n, \mathbb{R})$ denote the set of n by n real matrices. It is also a vector space. In fact, $M(n, \mathbb{R})$ is isomorphic to the direct sum of $\mathbb{R}^n$ with itself n times since we may write matrices in $M(n, \mathbb{R})$ as n-tuples $A = (A_1, \ldots, A_n)$, where $A_i \in \mathbb{R}^n$ is a column vector. In other words, $M(n, \mathbb{R})$ is isomorphic to $\mathbb{R}^{n^2}$. Of course, $M(n, \mathbb{R})$ is more than a vector space: it is also a ring with product given by matrix multiplication.

An element $A \in M(n, \mathbb{R})$ may be regarded as a linear map $A : \mathbb{R}^n \to \mathbb{R}^n$ and the norm of A is the (finite) number

$$|A| := \max\{|Av| : v \in \mathbb{R}^n, |v| = 1\}.$$

We also need to consider *multilinear maps.* A k-linear map on $\mathbb{R}^n$ with values in $\mathbb{R}^m$ is a function of the form

$$B : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}^m$$

which is linear in each of the k arguments. The smallest C for which

$$|B(v_1, \cdots, v_k)| \leq C|v_1| \cdots |v_k|$$

for all $v_1, \ldots, v_k$, is by definition the *norm of* B, denoted $|B|$. To see that $|B|$ is finite, set $w_i = v_i/|v_i|$, and note that $B(w_1, \cdots, w_k)$ must be bounded since it is a continuous function on the compact set

$$\{(w_1, \ldots, w_2) \in \mathbb{R}^l \times \cdots \times \mathbb{R}^l \ : \ |w_1| = \cdots = |w_k| = 1\}.$$

Definition 3.1.1 (Differentiable function). Let $U \subset \mathbb{R}^n$ be an open set. A function $f : U \to \mathbb{R}^m$ is said to be *differentiable* at $p \in U$ if there exists a linear map $df_p : \mathbb{R}^n \to \mathbb{R}^m$ that approximates f near p in the following sense: for all v such that $p + v \in U$,

$$f(p + v) = f(p) + df_p v + r(p, v)$$

where the function $r(p, v)$ satisfies

$$\lim_{v \to 0} r(p, v)/|v| = 0$$

for each p.

The linear approximation df_p, when it exists, is called the *differential* of f at p. It is immediate from the definition that

$$df_pv = \lim_{t\to 0} \frac{f(p+tv)-f(p)}{t} = \left[\frac{d}{dt}f(p+tv)\right]_{t=0}.$$

Note that $p + tv$ describes a parametric line through p with direction v, so df_pv is obtained by restricting f to that line and differentiating with respect to the parameter.

Denote by x_i the i-th coordinate function of $\mathbb{R}^n$. It is the function on $\mathbb{R}^n$ that associates to each $p = (a_1, \ldots, a_n)$ the i-th coordinate $x_i(p) = a_i$. Write $f(p) = (f_1(p), \ldots, f_m(p))$. The partial derivative of f_j at p with respect to x_i will on different occations be written in a number of different ways:

$$D_{e_i}f_j = \partial_i f_j = \frac{\partial f_j}{\partial x_i}.$$

We can express df_p as a matrix by using the standard basis of $\mathbb{R}^n$ and $\mathbb{R}^m$, as follows. Note that

$$df_pe_i = \sum_{j=1}^m \left(\lim_{t\to 0}\frac{f_j(p+te_i)-f_j(p)}{t}\right)e_j = \sum_{j=1}^m \frac{\partial f_j}{\partial x_i}(p)e_j.$$

Therefore, the matrix of df_p has (i,j)-entry $\partial_j f_i(p)$. If $v = a_1e_1 + \cdots + a_ne_n$, then $df_pv = b_1e_1 + \cdots + b_me_m$, where

$$\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} = \begin{pmatrix} \partial_1 f_1 & \partial_2 f_1 & \cdots & \partial_n f_1 \\ \partial_1 f_2 & \partial_2 f_2 & \cdots & \partial_n f_2 \\ \cdots & \cdots & \cdots & \cdots \\ \partial_1 f_m & \partial_2 f_m & \cdots & \partial_n f_m \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

and all derivatives are calculated at p. The above matrix of partial derivatives is called the *Jacobian matrix* of f (at p). We denote it by $Jf(p)$.

EXERCISE 3.1.2. If f and g are functions into $\mathbb{R}$, both differentiable at $p \in \mathbb{R}^m$, show that fg is differentiable at p and its differential satisfies

$$d(fg)_p = g(p)df_p + f(p)dg_p.$$

a. Affine Maps. As a first example, consider an *affine map* $f : \mathbb{R}^n \to \mathbb{R}^m$, which is a function of the form $f(p) = c + Tp$, where $c \in \mathbb{R}^m$ is constant and $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map. Then

$$f(p+v) - f(p) = Tv$$

so f is differentiable, with the remainder function r equal to 0. Furthermore $df_p = T$ for all $p \in \mathbb{R}^n$. In particular, the differential of a linear map T at any point is T itself.

b. Bilinear Maps. Consider a bilinear map $B : \mathbb{R}^l \times \mathbb{R}^l \to \mathbb{R}^m$. It is differentiable at each point $p = (p_1, p_2)$ of $\mathbb{R}^n = \mathbb{R}^l \times \mathbb{R}^l$, and the differential at p can be written as follows. Let $v = (v_1, v_2) \in \mathbb{R}^l \times \mathbb{R}^l$. Then I claim that

$$dB_pv = B(p_1, v_2) + B(v_1, p_2)$$

which is clearly linear in v. To verify this identity, first note that

$$B(p_1 + v_1, p_2 + v_2) - B(p_1, p_2) = B(p_1, v_2) + B(v_1, p_2) + B(v_1, v_2)$$

and that

$$\frac{|B(v_1,v_2)|}{|(v_1,v_2)|} = \frac{|B(v_1,v_2)|}{(|v_1|^2+|v_2|^2)^{\frac{1}{2}}} \le |B|\frac{|v_1||v_2|}{(|v_1|^2+|v_2|^2)^{\frac{1}{2}}} \le |B|(|v_1|^2+|v_2|^2)^{\frac{1}{2}}/2.$$

Therefore, taking $r(p,v) = B(v_1,v_2)$, we have $|r(p,v)|/|v| \le |B||v|/2$, which obviously approaches 0 as $v \to 0$. If $Q : \mathbb{R}^l \to \mathbb{R}^n$ is the quadratic map defined by $Q(p) = B(p,p)$, where B is a symmetric bilinear map, then it follows that $dQ_p v = 2B(p,v)$.

EXERCISE 3.1.3. Find the differential of a k-linear map $Q(p) := L(p,\dots,p)$, where $L : \mathbb{R}^l \times \cdots \times \mathbb{R}^l \to \mathbb{R}^n$. is a k-linear symmetric map. ('Symmetric' means that $L(p_1,\dots,p_k)$ does not change under any permutation of the $p_1,\dots,p_k$.)

EXERCISE 3.1.4. Show that the function $f : M(n,\mathbb{R}) \to M(n,\mathbb{R})$ given by $f(X) = X^3$ is differentiable at all points. Find the differential df_A at a point A.

c. Differentiable Paths. A *path*, or *curve*, in $\mathbb{R}^n$ is a map c from an interval $I \subset \mathbb{R}$ into $\mathbb{R}^n$. If c is differentiable at an interior point t_0 of I, the *velocity vector* at t_0 is

$$c'(t_0) := \lim_{h\to 0} \frac{c(t_0+h)-c(t_0)}{h}.$$

We think of $c'(t_0)$ as a tangent vector to the path at $c(t_0)$. For example, if $c(t) = a + bt$ is the straight line that starts at $a \in \mathbb{R}^n$ for $t = 0$ and is parallel to $b \in \mathbb{R}$, we have $c'(t) = b$ for all t.

Let e denote the unit basis vector on $\mathbb{R}$. The velocity vector $c'(t)$ can be written as $dc_t e$, i.e., the image of e under the differential of c at $t \in I$. If $c(t) = (f_1(t),\dots,f_n(t))^t$, then $c'(t) = \sum_{i=1}^n f_i'(t)e_i$.

EXERCISE 3.1.5. Given $c(t) = (t,t^2,\dots,t^n)$, find a parametric equation of the straight line through the point $p = (1,1,\dots,1) \in \mathbb{R}^n$ having the same velocity vector at p as c.

d. Functions into $\mathbb{R}$. We consider now the special case of functions from $\mathbb{R}^m$ into $\mathbb{R}$. Let U be a nonempty open set in $\mathbb{R}^n$ and let $f : U \to \mathbb{R}$ be differentiable at $p \in U$. If $v = (a_1,\dots,a_n)^t \in \mathbb{R}^n$, we have

$$df_p v = df_p \sum_{i=1}^n a_i e_i = \sum_{i=1}^n a_i df_p e_i = \sum_{i=1}^n a_i \frac{\partial f}{\partial x_i}(p).$$

The differential df_p is nothing but the ordinary differential defined in calculus. Recall that $x_i : \mathbb{R}^n \to \mathbb{R}$ denotes the i-th coordinate function on $\mathbb{R}^n$. It follows from the above that $dx_i(e_j) = 0$ if $i \ne j$ and $dx_i(e_i) = 1$. Therefore, denoting by $(\mathbb{R}^n)^*$ the dual vector space to $\mathbb{R}^n$ (which is, of course, isomorphic to $\mathbb{R}^n$ itself), then $\{dx_1,\dots,dx_n\}$ is a basis of $(\mathbb{R}^n)^*$ dual to $\{e_1,\dots,e_n\}$. Note that $dx_i(v) = a_i$, so $df_p v = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(p) dx_i(v)$. Therefore,

$$df = \sum_{i=1}^n \frac{\partial f}{\partial x_i} dx_i.$$

If $v = a_1e_1 + \cdots + a_ne_n$ is a vector of unit length, then $df_p v = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(p)a_i$ is what in calculus is called the *directional derivative* of f at p on the direction of v. Later we will also use the notation $(D_v f)(p)$ for the directional derivative.

e. The Chain Rule. If two composable maps f and g are differentiable then their composition $g \circ f$ is also differentiable and its differential is the composition of the differentials df and dg. This is the content of the chain rule. More precisely, we have:

PROPOSITION 3.1.6 (Chain rule). Let $U \subset \mathbb{R}^m$ and $V \subset \mathbb{R}^n$ be nonempty open sets, $f : U \to \mathbb{R}^n$ a differentiable function at $p \in U$ such that $f(U) \subset V$ and $g : V \to \mathbb{R}^k$ a differentiable map at $q = f(p) \in V$. Then $g \circ f : U \to \mathbb{R}^k$ is differentiable at p and $d(g \circ f)_p = dg_q \circ df_p$.

PROOF. We can find functions $\sigma_1 : \mathbb{R}^m \to \mathbb{R}^n$ and $\sigma_2 : \mathbb{R}^n \to \mathbb{R}^k$ such that $\sigma_i(h_i)/|h_i|$ goes to 0 as h_i goes to 0 and

$$f(p + h_1) = f(p) + df_p h_1 + \sigma_1(h_1), \quad g(q + h_2) = g(q) + dg_q h_2 + \sigma_2(h_2).$$

Therefore

$$(g \circ f)(p + h) = g(f(p)) + dg_{f(p)} df_p h + \tau(h)$$

where $\tau(h) = dg_{f(p)}\sigma_1(h) + \sigma_2(k)$ and $k = df_p h + \sigma_1(h)$. It is now immediate that $\tau(h)/|h|$ goes to 0 as h goes to 0 so that $f \circ g$ is differentiable and has the claimed differential. □

EXERCISE 3.1.7. Show that the Jacobian matrices of f, g and $g \circ f$ satisfy

$$J(g \circ f)(p) = Jg(f(p))Jf(p)$$

i.e.,

$$\frac{\partial(g_i \circ f)}{\partial x_j}(p) = \sum_{k=1}^{n} \frac{\partial g_i}{\partial y_k}(f(p))\frac{\partial f_k}{\partial x_j}(p).$$

COROLLARY 3.1.8. Let $f : U \subset \mathbb{R}^m \to \mathbb{R}^n$ be differentiable at $p \in U$. Let $v \in \mathbb{R}^m$ and let $c : (-\epsilon, \epsilon) \to U$ be a path differentiable at 0 such that $c(0) = p$ and $c'(0) = v$. Then $df_p v$ is the velocity vector of the path $t \mapsto f(c(t))$ at $t = 0$.

PROOF. By the chain rule, $(f \circ c)'(0) = df_p c'(0) = df_p v$. □

f. Directional Derivatives. The dot product of two vectors $u, v \in \mathbb{R}^m$ will be written $\langle u, v\rangle$. Let $f : U \to \mathbb{R}$ be a differentiable function defined on an open subset U of $\mathbb{R}^m$. The *gradient* of f at $p \in U$ is the unique vector u characterized by the property

$$\langle u, h\rangle = df_p h.$$

The gradient of f at p will be denoted $\mathrm{grad}(f)_p$.

A *level set* of f is a set of the form $f^{-1}(c)$.

EXERCISE 3.1.9. Let $f : U \subset \mathbb{R}^m \to \mathbb{R}$ be a differentiable function on an open subset $U \subset \mathbb{R}^m$.

(1) If a differentiable path $c : (a, b) \to \mathbb{R}^m$ maps into a level set of f, show that $c'(t)$ is perpendicular to $\mathrm{grad}(f)_{c(t)}$.
(2) Show that for each $p \in U$, the maximum value of $df_p v$ for $|v| = 1$ is $|\mathrm{grad}(f)_p|$ and it is achieved exactly when v is a positive multiple of the gradient of f at p.

It will be convenient to set the following notation. If v is a vector in $\mathbb{R}^n$ and $f : U \subset \mathbb{R}^m \to \mathbb{R}$ is a function that is differentiable at $p \in U$, then we write

$$v_p f := df_p v.$$

Since $d(fg)_p = g(p)df_p + f(p)dg_p$ for two functions that are differentiable at p, we have

$$v_p(fg) = (v_p f)g(p) + f(p)v_p g.$$

This shows that v_p satisfies the product rule for derivations. Indeed, if we write $v = a_1e_1 + \cdots + a_me_m$, then as we already know, $df_p v = \sum_{i=1}^m a_i \frac{\partial f}{\partial x_i}(p)$, so that v_p is the derivation

$$v_p = \sum_{i=1}^m a_i (\frac{\partial}{\partial x_i})_p.$$

Under this correspondence between vectors and derivations, the standard basis vectors of $\mathbb{R}^m$, e_i, are associated to the partial derivatives $(\frac{\partial}{\partial x_i})_p$. We will refer to v_p as a *vector at* p, and freely use the notation $v_p f$ for $df_p v$ when it is more convenient. Sometimes $(vf)(p)$ will also be used.

When a vector is regarded as a derivation, as above, it is naturally attached to a point in $\mathbb{R}^n$, namely, the point where the derivative is evaluated. In this sense we talk about a *vector at a point* p. The set of vectors at p forms a vector space naturally isomorphic to $\mathbb{R}^n$, which we will denote by $\mathbb{R}^n_p$, or sometimes $T_p\mathbb{R}^n$. This space will be called the *tangent space to* $\mathbb{R}^n$ *at* p. Note that, from this point of view, the differential df_p of a differentiable map $f : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map from $\mathbb{R}^n_p$ to $\mathbb{R}^m_{f(p)}$. In fact, if $v = \gamma'(0)$ for a differentiable curve $\gamma(t)$ such that $\gamma(0) = p$, we have seen by corollary 1.1.8 that $df_p v = (f \circ \gamma)'(0)$ hence, viewed as a derivation, $df_p v$ should be regarded as a derivative at $f(p)$. If $h : \mathbb{R}^m \to \mathbb{R}$ is a differentiable function at $f(p)$, we obtain by the chain rule:

$$(df_p v)h := dh_{f(p)} df_p v = d(h \circ f)_p v = v_p(h \circ f).$$

2. Differentials of higher order

Let $U \subset \mathbb{R}^m$ be a nonempty open set. We say that $f : U \to \mathbb{R}^n$ is differentiable on U if it is differentiable at each $p \in U$. The differential df_p is, for each p, an element of the space $L(\mathbb{R}^m, \mathbb{R}^n)$ of linear maps from $\mathbb{R}^m$ into $\mathbb{R}^n$. Therefore, if f is differentiable on U we obtain a map $p \in U \mapsto df_p \in L(\mathbb{R}^m, \mathbb{R}^n)$. For simplicity, we sometimes denote this map by f', so that $f'(p) := df_p$. We say that f is a function of *class* C^1 on U if f is differentiable over U and $f' : U \to L(\mathbb{R}^m, \mathbb{R}^n)$ is continuous.

Since $L(\mathbb{R}^m, \mathbb{R}^n)$ is isomorphic to $\mathbb{R}^{mn}$ it makes sense to ask whether f' is itself differentiable. If f' is differentiable at p we say that f is twice differentiable at p and denote its differential by either one of the following notations:

$$f''(p) = d_2 f_p.$$

Note that $f''(p)$ is a linear map from $\mathbb{R}^m$ into $L(\mathbb{R}^m, \mathbb{R}^n)$, hence an element of $L(\mathbb{R}^m, L(\mathbb{R}^m, \mathbb{R}^n))$. If f is twice differentiable at all $p \in U$ we say that f is twice differentiable on U. If moreover $f'' : U \to L(\mathbb{R}^m, L(\mathbb{R}^m, \mathbb{R}^n))$ is continuous we say that f is of *class* C^2.

Denote by $L_2(\mathbb{R}^m, \mathbb{R}^n)$ the space of bilinear maps from $\mathbb{R}^m \times \mathbb{R}^m$ into $\mathbb{R}^n$. There is a natural isomorphism between the spaces $L(\mathbb{R}^m, L(\mathbb{R}^m, \mathbb{R}^n))$ and $L_2(\mathbb{R}^m, \mathbb{R}^n)$.

In fact, for each $T \in L(\mathbb{R}^m, L(\mathbb{R}^m, \mathbb{R}^n))$, define $\bar{T} \in L_2(\mathbb{R}^m, \mathbb{R}^n)$ by

$$\bar{T}(u, v) := (Tu)v.$$

Under this isomorphism the second order differential of f at p may be regarded as a bilinear map

$$d_2 f_p : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^n.$$

We saw before that $(dx_1, \ldots, dx_m)$ is the dual basis of $(e_1, \ldots, e_m)$ and that $f'(p) = \sum_{i=1}^m \frac{\partial f}{\partial x_i}(p) dx_i$.

EXERCISE 3.2.1. Show that $d_2 f_p(e_i, e_j) = \frac{\partial^2 f}{\partial x_i \partial x_j}(p)$.

The space $L_2(\mathbb{R}^m, \mathbb{R})$ has a basis denoted by $dx_i \otimes dx_j$, which is defined by

$$dx_i \otimes dx_j(u, v) := dx_i(u) dx_j(v).$$

Therefore,

$$d_2 f = \sum_{i,j=1}^m \frac{\partial^2 f}{\partial x_i \partial x_j}(p) dx_i \otimes dx_j.$$

Higher order differentials are defined inductively. We write

$$f^{(k)}(p) := d_k f(p)$$

for the k-th order differential, which is an element of the vector space of k-linear maps from $\mathbb{R}^m \times \cdots \times \mathbb{R}^m$ into $\mathbb{R}^n$, denoted here by $L_k(\mathbb{R}^m, \mathbb{R}^n)$. If $d^k f_p$ exists, we say that f is k times differentiable at p. If f is k times differentiable at each point of U we say that f is k times differentiable on U. If, in this case, $f^{(k)}$ is also continuous on U we say that f is of *class* C^k on U, or a C^k function on U. We also write $f \in C^k$. If f belongs to the intersection of all C^k, we write $f \in C^\infty$, and say that f is of class C^∞, or that f is a C^∞ function on U. C^∞ functions are also called *smooth*.

The elements $dx_{i_1} \otimes \cdots \otimes dx_{i_k} \otimes e_j \in L_k(\mathbb{R}^m, \mathbb{R}^n)$ defined by

$$(dx_{i_1} \otimes \cdots \otimes dx_{i_k} \otimes e_j)(v_1, \ldots, v_k) = dx_{i_1}(v_1) \cdots dx_{i_k}(v_k) e_j$$

form a basis of $L_k(\mathbb{R}^m, \mathbb{R}^n)$.

PROPOSITION 3.2.2. If f is a C^k function on U, then $d_k f$ is a k-linear symmetric map. In other words,

$$d_k f_p(u_1, \ldots, u_k) = d_k f_p(u_{\sigma(1)}, \ldots, u_{\sigma(k)})$$

for any permutation σ of $\{1, \ldots, k\}$.

PROOF. Once one knows the above for $k = 2$, the general cases follows by a simple induction. The case $k = 2$ is an elementary but tedious calculation and will be omitted. □

Note that

$$d_k f_p(e_{i_1}, \ldots, e_{i_k}) = \frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}}(p)$$

so the previous proposition is simply an expression of the well known fact that if all derivatives of f up to order k are continuous then the order of partial derivatives is immaterial.

EXERCISE 3.2.3. Let $A : \mathbb{R}^m \times \cdots \times \mathbb{R}^m \to \mathbb{R}^n$ be a k-linear map. Define $f : \mathbb{R}^m \to \mathbb{R}^n$ by $f(x) = (1/n!)A(x, \ldots, x)$. Find all the differentials $d_i f_x$.

EXERCISE 3.2.4. Let $f : \mathbb{R}^m \to \mathbb{R}^n$ be a smooth function such that $f(tx) = t^k f(x)$ for all $t \in \mathbb{R}$ and $x \in \mathbb{R}^m$. We say that f is a *homogeneous function of degree* k. Show that for each i, $1 \le i \le k$,

$$d_i f_x(h_1, \dots, h_i) = \frac{1}{(k-i)!} d_k f_0(x, \dots, x, h_1, \dots, h_i)$$

and all the differentials of order higher than k are 0. (Hint: differentiate i times $f(tx) = t^k f(x)$ with respect to x and show that $f^{(i)}(x) := d^i f_x$ is homogeneous of degree $k - i$. Then differentiate $k - i$ times $f^{(i)}(tx) = t^{k-i} f^{(i)}(x)$ with respect to t.)

We state without proof Taylor's formula.

PROPOSITION 3.2.5 (Taylor's formula). Let $U \subset \mathbb{R}^m$ be a nonempty open set and $f : U \to \mathbb{R}$ a k times differentiable function on U such that $d_{k+1} f_p$ exists for a $p \in U$. Then

$$f(p+h) = f(p) + df_p h + \frac{1}{2!} d_2 f_p(h,h) + \cdots + \frac{1}{(k+1)!} d_{k+1} f_p(h, \dots, h) + \sigma(h)$$

where $\sigma(h)/|h|^{k+1}$ goes to zero with h. Moreover,

$$\sigma(h) = \int_0^1 \frac{(1-t)^k}{k!} (d_{k+1} f)_{p+th}(h, \dots, h)\, dt.$$

3. Local Properties of Differentiable Maps

Let $U, V \subset \mathbb{R}^m$ be open. A differentiable bijection $f : U \to V$ whose inverse is also differentiable is called a *diffeomorphism* from U to V. If both f and its inverse f^{-1} are C^k we say that f is a C^k diffeomorphism.

For example, $f : (0, \infty) \to (0, \infty)$ given by $f(x) = x^3$ is a diffeomorphism. The map $f : \mathbb{R} \to \mathbb{R}$ given by by the same expression, $f(x) = x^3$, is a C^∞ bijection but not a diffeomorphism since the inverse, $f^{-1}(x) = x^{1/3}$, is not differentiable at 0.

We say that $f : U \to \mathbb{R}^m$ is a *local diffeomorphism* if for each $p \in U$ there are open neighborhoods $V \subset U$ of p and W of $f(p)$ such that the restriction $f|_V : V \to W$ is a diffeomorphism.

EXERCISE 3.3.1. Show that $f : \mathbb{R}^2 \to \mathbb{R}^2 - \{0\}$ defined by

$$f(x, y) = (e^x \cos y, e^x \sin y)$$

is a local C^∞ diffeomorphism.

a. The Inverse Function Theorem. The next theorem is one of the most fundamental in calculus.

THEOREM 3.3.2 (Inverse function). Let $U \subset \mathbb{R}^m$ be open and $f : U \to \mathbb{R}^m$ be C^k, $1 \le k \le \infty$, such that at some $p \in U$ the differential df_p is a linear isomorphism. Then f is a C^k diffeomorphism from a neighborhood V of p onto a neighborhood W of $f(p)$.

PROOF. Only a rough sketch of the proof will be given here. There is no loss of generality in assuming that $p = 0$ and $f(0) = 0$. Since f is differentiable at 0, we can also assume, by taking U small enough, that $f(q) = Tq + \Phi(q)$ for $q \in U$, where $T \in GL(m, \mathbb{R})$ and $\Phi : U \to \mathbb{R}^m$ satisfies $|\Phi(q_1) - \Phi(q_2)| \le \lambda |q_1 - q_2|$ for some λ such that $\lambda |T^{-1}| < 1$.

Write $(T^{-1} \circ f)(q) = q + (T^{-1} \circ \Phi)(q)$. The map $\eta = T^{-1} \circ \Phi$ can be shown to be a *contraction* near 0. This means that

$$|\eta(q_1) - \eta(q_2)| \leq \lambda |q_1 - q_2|$$

for all q_1 and q_2 in a neighborhood of 0, for some $\lambda < 1$. Therefore, $\Psi := T^{-1} \circ f$ is a perturbation of the identity function by adding to it a contraction. Such a perturbation of the identity must be one-to-one since

$$\begin{aligned} |\Psi(q_1) - \Psi(q_2)| &= |q_1 - q_2 + \eta(q_1) - \eta(q_2)| \\ &\geq |q_1 - q_2| - |\eta(q_1) - \eta(q_2)| \\ &\geq (1 - \lambda)|q_1 - q_2|, \end{aligned}$$

whence it also follows that the inverse map is continuous. Some extra work using a well-known theorem to the effect that any contraction of a complete metric space must have a unique fixed point gives that the perturbation of the identity is also an open map.

Note that f differs from the perturbation of the identity by the diffeomorphism T^{-1}. The upshot is that we can find an inverse homeomorphism $g = f^{-1} : W \to V$, where V and W are open neighborhoods of 0.

It remains to show that g is differentiable at each point $f(q)$. It is clear that if it is, then by the chain rule the differential of g must be $(df_q)^{-1}$. Write $u = f(q)$ and define a function s by the expression

$$g(u + h) = g(u) + (df_q)^{-1} h + s(h).$$

It is now a calculation to check that $s(h)/|h|$ indeed goes to 0 with h, but this is precisely what is needed to prove differentiability. □

COROLLARY 3.3.3. Let $U \subset \mathbb{R}^m$ be open and $f : U \to \mathbb{R}^m$ a C^k map, $k \geq 1$. Then f is a C^k local diffeomorphism if and only for each $p \in U$, df_p is an isomorphism.

EXERCISE 3.3.4. Denote by S the space of all symmetric m by m matrices all of whose eigenvalues are positive. (In other words, S is the space of *positive definite* symmetric matrices in $M(m, \mathbb{R})$.) Show that $f(X) = X^2$ is a diffeomorphism from S onto itself.

b. Submersions. We now discuss some consequences of the inverse function theorem. Let $U \subset \mathbb{R}^m$ be a nonempty open set. A differentiable map $f : U \to \mathbb{R}^n$ is called a *submersion* if for every $p \in U$ the differential $df_p : \mathbb{R}^m \to \mathbb{R}^n$ is surjective. Of course, this is only possible for $m \geq n$. Also note that if f is C^1 and df_p is surjective for some p then df_q is surjective for all q in some sufficiently small neighborhood of p. This is a consequence of the following exercise.

EXERCISE 3.3.5. Let $T_q : \mathbb{R}^m \to \mathbb{R}^n$ be a family of linear maps depending continuously on $q \in U$, where U is an open subset of $\mathbb{R}^k$. Suppose that T_p is surjective. Show that T_q is surjective for all q in some neighborhood of p. (Note that T_q is an element of $L(\mathbb{R}^m, \mathbb{R}^n) = \mathbb{R}^{mn}$. So the family T is a function from U into $\mathbb{R}^{mn}$, and continuity makes sense.)

For example, the projection $P : \mathbb{R}^{m+n} = \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ defined by $P(x, y) = y$ is a submersion at every point. The next theorem says that after a change of coordinates (given in the theorem by the diffeomorphism h) any submersion locally looks like P.

THEOREM 3.3.6 (Local form of submersions). Let $U \subset \mathbb{R}^{m+n}$ be an open set and $f : U \to \mathbb{R}^n$ a C^k function, $k \geq 1$. Suppose that at $p \in U$ the differential $df_p : \mathbb{R}^{m+n} \to \mathbb{R}^n$ is surjective. Let $\mathbb{R}^{m+n} = E_1 \oplus E_2$ be any direct sum decomposition, with $p = (p_1, p_2)$, such that $df_p|_{E_2}$ is an isomorphism from E_2 onto $\mathbb{R}^n$. Then there exist open neighborhoods $V \subset E_1$ of p_1, $W \subset \mathbb{R}^n$ of $f(p)$ and $U' \subset U$ of p, and a C^k diffeomorphism $h : V \times W \to U'$ such that $f \circ h(q, u) = u$ for all $(q, u) \in V \times U'$.

PROOF. Define a C^k map $\Phi : U \to E_1 \times \mathbb{R}^n$ by setting

$$\Phi(x, y) = (x, f(x, y)).$$

We write $\partial_i f_p := df_p|_{E_i}$. The differential $d\Phi_p : \mathbb{R}^{m+n} = E_1 \oplus E_2 \to E_1 \oplus \mathbb{R}^n$ is given by

$$(h_1, h_2) \mapsto (h_1, (\partial_1 f_p)h_1 + (\partial_2 f_p)h_2).$$

We claim that $d\Phi_p$ is an isomorphism. In fact, the map

$$(u_1, u_2) \mapsto (u_1, [\partial_2 f_p]^{-1}(u_2 - (\partial_1 f_p)h_2))$$

is an inverse for $d\Phi_p$ as one easily checks. By the inverse function theorem Φ is a local C^k diffeomorphism from a neighborhood of p onto a neighborhood of $(p_1, f(p))$. The latter neighborhood can be chosen of the form $V \times W$, where V is open in E_1 and W is open in $\mathbb{R}^n$.

Set $U' = \Phi^{-1}(V \times W)$ and $h = \Phi^{-1} : V \times W \to U'$. Note that $f \circ h(x, w) = w$ for every (x, w) in $V \times W$. In fact, since $\Phi(x, y) = (x, f(x, y))$ it follows that h has the form $h(x, w) = (x, H(x, w))$. Therefore for every $(x, w) \in V \times W$ we have $(x, w) = \Phi(h(x, w)) = (x, f \circ h(x, w))$ from which it follows that $w = f \circ h(x, w)$. □

COROLLARY 3.3.7. A C^k submersion, for $k \geq 1$, is an open map.

PROOF. The projection P is certainly an open map and so is a diffeomorphism h. Therefore the submersion, which is the composition of the two open maps, is also an open map. □

The expressions $\partial_i f$ introduced in the previous proof will be called the *partial differentials* of f at p with respect to the direct sum decomposition $\mathbb{R}^{m+n} = E_1 \oplus E_2$.

THEOREM 3.3.8 (Implicit function). Let U be an open set in $\mathbb{R}^{m+n}$ and let $f : U \to \mathbb{R}^n$ be a C^k map, $k \geq 1$. Suppose that $\mathbb{R}^{m+n} = E_1 \oplus E_2$ is a direct sum decomposition such that, for $p = (p_1, p_2) \in U$, the second partial differential $\partial_2 f_p E_2 \to \mathbb{R}^n$ is an isomorphism. Set $c = f(p)$. Then there exist open sets $V \subset E_1$ containing p_1 and $U' \subset U$ containing p such that for each $x \in V$ there is a unique $g(x) \in E_2$ for which $(x, g(x)) \in U'$ and $f(x, g(x)) = c$. The map g is C^k and its differential is given by

$$dg_x = -[\partial_2 f_{(x,g(x))}]^{-1} \circ \partial_1 f_{(x,g(x))}.$$

PROOF. Note that f is a submersion at p. We use here the same notations as in the proof of the previous theorem. In particular, we set $U' = h(V \times W)$ and $h(x, y) = (x, H(x, y))$ for all $(x, y) \in V \times W$. Define a C^k function $g : V \to E_2$ by $g(x) = H(x, c)$. Then $f(x, g(x)) = f(h(x, c)) = c$ for all $x \in V$. If $(x, u) \in U'$ also satisfies $f(x, u) = c$ then, for Φ also as in the proof of the previous theorem,

$$(x, u) = h(\Phi(x, u)) = h(x, c) = (x, H(x, c)) = (x, g(x))$$

so that $u = g(x)$ and $g(x)$ is unique. Differentiating $f(x, g(x)) = c$ we obtain

$$\partial_1 f_{(x,g(x))} + \partial_2 f_{(x,g(x))} \circ dg_x = 0$$

so that the differential of g is as claimed. $\square$

c. Immersions. A differentiable map $f : U \to \mathbb{R}^n$ from an open subset $U \subset \mathbb{R}^m$ is called an *immersion* if for every $p \in U$ the differential $df_p : \mathbb{R}^m \to \mathbb{R}^n$ is injective. This requires $m \leq n$. The main example is the inclusion map $i : \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^{n-m}$ given by $i(x) = (x, 0)$.

EXERCISE 3.3.9. Let $T_q : \mathbb{R}^m \to \mathbb{R}^{m+n}$ be a family of linear maps parametrized by $q \in U \subset \mathbb{R}^k$ such that $q \mapsto T_q$ is continuous. If T_p is injective for some $p \in U$, show that T_q is injective for all q in some neighborhood of p.

The next theorem shows that any immersion locally looks like i.

THEOREM 3.3.10 (Local form of immersions). Let $U \subset \mathbb{R}^m$ be an open set and $f : U \to \mathbb{R}^{m+n}$ a C^k map, $k \geq 1$. Suppose that at $p \in U$ the differential df_p is injective. Then there exist neighborhoods Z of $f(p)$, V of p, and $W \subset \mathbb{R}^n$ of 0, and a C^k diffeomorphism $h : Z \to V \times W$ such that $h(f(x)) = (x, 0)$.

PROOF. Set $E_1 = df_p\mathbb{R}^m$ and let E_2 be any complementary subspace in $\mathbb{R}^{m+n}$, so that $\mathbb{R}^{m+n} = E_1 \oplus E_2$ is a direct sum decomposition. Note, in particular, that $df_p : \mathbb{R}^m \to E_1$ is a linear isomorphism. Define $\Phi : U \times E_2 \to \mathbb{R}^{m+n}$ by $\Phi(x, y) = f(x) + y$. Then Φ is C^k, $\Phi(p, 0) = f(p)$. Moreover, for any $(u, v) \in \mathbb{R}^m \oplus E_2$ we have $d\Phi_{(p,0)}(u, v) = df_p u + v$. Therefore, $d\Phi_{(p,0)}$ is an isomorphism from $\mathbb{R}^m \oplus E_2$ onto $\mathbb{R}^{m+n}$. By the inverse function theorem, Φ is a C^k diffeomorphism from a neighborhood of $(p, 0)$, which can be chosen to be of the form $V \times W$, where V is a neighborhood of p in U and W is a neighborhood of 0 in E_2, onto a neighborhood Z of $f(p)$ in $\mathbb{R}^{m+n}$. Let h be the inverse to $\Phi|_{V \times W}$. Since $\Phi(x, 0) = f(x)$, we have $h(f(x)) = h(\Phi(x, 0)) = (x, 0)$, for all $x \in V$. We can now identify E_2 with $\mathbb{R}^n$ by a choice of basis, concluding the proof. $\square$

COROLLARY 3.3.11. Let $f : U \to \mathbb{R}^{m+n}$ be a C^k map such that $df_p : \mathbb{R}^m \to \mathbb{R}^{m+n}$ is injective for some $p \in U$. Then there exists a neighborhood V of p such that $f : V \to f(V)$ is a homeomorphism whose inverse $f^{-1} : f(V) \to V$ is the restriction of a C^k map $g : Z \to V$, where Z is a neighborhood of $f(p)$.

PROOF. It suffices to take V and Z as in the theorem and define $g = \Pi \circ h$, where $\Pi : V \times W \to V$ is the first projection. Then for $x \in V$ we have $g(f(x)) = \Pi(h(f(x))) = \Pi(x, 0) = x$, so that the restriction of g to $f(V)$ is f^{-1}. Since g is clearly C^k the corollary follows. $\square$

d. The Rank Theorem. We state next a very useful theorem, which generalizes the theorems on the local form of both submersions and immersions, called the *rank theorem*.

Recall that the *rank* of a linear map $T : \mathbb{R}^m \to \mathbb{R}^n$ is the dimension of the image $T(\mathbb{R}^m)$. Equivalently, it is the maximal number of linearly independent vectors among $Te_1, \ldots, Te_m$. Thus, the rank of T is r if, relative to a choice of basis, the matrix of T has a minor determinant of order r not equal to 0, and all minor determinants of order $r + 1$ are 0.

For a differentiable map $f : U \subset \mathbb{R}^m \to \mathbb{R}^n$ the rank at a point $p \in U$ is by definition the rank of the differential of f at p. For example, if $f : U \to \mathbb{R}^n$ is a submersion, it has rank n at all points in U, whereas if it is an immersion then the rank is m at all points in U. Note that if $f : U \to \mathbb{R}^n$ has the maximum rank allowed by the dimensions n, m, then it must be either a submersion or an immersion.

If the rank of f is not maximum at a point $p \in U$ there may not be any neighborhood of p where rank is constant, so the local form of f near p cannot have a simple description as for submersions and immersions. For example, the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x, y) = x^2 + y^2$ is not a submersion at $(0, 0)$ and near $(0, 0)$ f does not look like projections or inclusions, or any combination of these maps. (The level sets of f are concentric circles collapsing to the origin.) However, if we assume that the rank of f is constant on some open set, then f will locally look like a projection composed with an inclusion. This is the content of the rank theorem.

THEOREM 3.3.12 (Rank theorem). Let U be an open set in $\mathbb{R}^{m+n}$ and let $f : U \to \mathbb{R}^{m+p}$ be a C^k map, $k \geq 1$. Assume that the rank of f is constant on U and equal to m. Then, for every $p \in U$ there exist C^k diffeomorphisms α from an open subset of $\mathbb{R}^m \times \mathbb{R}^n$ onto a neighborhood of p, and β from a neighborhood of $f(p)$ onto an open subset of $\mathbb{R}^m \times \mathbb{R}^p$ such that $(\beta \circ f \circ \alpha)(x, y) = (x, 0)$.

PROOF. The main ingredients are the theorems on local forms of immersions and submersions. Since there are no new ideas beyond those used to prove the latter theorems, we omit the details. It is left to the reader to interpret the theorem in the particular case for which f is a linear transformation, and to imagine how the previous results of the section can be applied here. □

PROPOSITION 3.3.13. Let $f : U \to \mathbb{R}^n$ be a C^1 function defined on an open subset $U \subset \mathbb{R}^m$. For each $r = 1, 2, \ldots, p$, for $p = \min\{m, n\}$, let A_r denote the interior of the set of $x \in U$ at which f has rank r (hence an open set). Then

$$A = A_0 \cup A_1 \cup \cdots \cup A_p$$

is an open and dense subset of U.

PROOF. Let V be any nonempty open subset of U and let $x \in V$ be a point at which the rank of $f|_V$ is maximum. There is a neighborhood W of x on which the rank of f is nondecreasing, so by taking W sufficiently small the rank of f is constant on W. This shows that the set where the rank of f is locally constant is an open dense set. Clearly the rank cannot be greater than $p = \min\{m, n\}$, so the proposition holds. □

4. Submanifolds of $\mathbb{R}^n$

A submanifold of dimension m in $\mathbb{R}^n$ is the m-dimensional analogye of a surface in $\mathbb{R}^3$. We know from calculus a number of ways in which surfaces in $\mathbb{R}^3$ can be represented: as the graph of a real valued function on the plane, as a level set of a real valued function on $\mathbb{R}^3$, or in parametric form. The definition of a submanifold of $\mathbb{R}^n$ will be based on the idea of local parametrizations.

Let U_0 be an open subset of $\mathbb{R}^m$. A C^k immersion $\varphi : U_0 \to \mathbb{R}^n$ is said to be a C^k *imbedding* in $\mathbb{R}^n$ if φ is a homeomorphism from U_0 onto $\varphi(U_0)$. In this case, we say that φ is a (m-dimensional) C^k *parametrization* of the subset $U = \varphi(U_0)$ of $\mathbb{R}^n$.

A 1-dimensional C^k parametrization $\varphi : J \to \mathbb{R}^n$, $(m = 1)$ describes a C^k path in $\mathbb{R}^n$. The set J is an open interval in $\mathbb{R}$. The condition that φ be an imbedding is equivalent to: $\varphi : J \to \varphi(J)$ is a homeomorphism and the velocity vector $\varphi'(t)$ is never 0. In particular, a path described by an imbedding φ cannot have self intersections and each image point $\varphi(t)$ has a neighborhood in $\mathbb{R}^n$ whose intersection with $\varphi(J)$ is a connected segment of curve.

Parametrizations of dimension 2 in $\mathbb{R}^3$ correspond to the familiar notion of a parametric surface studied in calculus. Let U_0 be an open subset of $\mathbb{R}^2$ and $\varphi : U_0 \to U = \varphi(U_0) \subset \mathbb{R}^3$ a parametrization of class C^k. Then φ can be written as follows:

$$\varphi(u,v) = (\varphi_1(u,v), \varphi_2(u,v), \varphi_3(u,v)).$$

The set U is called a *local surface*. The condition that φ be an immersion is equivalent to the linear independence of the two vectors

$$d\varphi_{(u,v)}e_1 = \frac{\partial \varphi}{\partial u}(u,v) = (\frac{\partial \varphi_1}{\partial u}(u,v), \frac{\partial \varphi_2}{\partial u}(u,v), \frac{\partial \varphi_3}{\partial u}(u,v))$$
$$d\varphi_{(u,v)}e_2 = \frac{\partial \varphi}{\partial v}(u,v) = (\frac{\partial \varphi_1}{\partial v}(u,v), \frac{\partial \varphi_2}{\partial v}(u,v), \frac{\partial \varphi_3}{\partial v}(u,v)).$$

As in calculus, we think of these vectors as being tangent to the local surface at $\varphi(u,v)$.

We now define an *m-dimensional (imbedded) submanifold of* $\mathbb{R}^n$ of class C^k as a nonempty subset $M \subset \mathbb{R}^n$ having the property that each $p \in M$ has a neighborhood $U \subset M$ (the intersection of an open subset of $\mathbb{R}^n$ with M) such that U admits a C^k parametrization of dimension m. The number $n - m$ is called the *codimension* of M in $\mathbb{R}^n$. A codimension one submanifold is also called a *hypersurface*.

A zero dimensional manifold is a set of isolated points. An n-dimensional submanifold of $\mathbb{R}^n$ is an open subset of $\mathbb{R}^n$.

a. The Sphere S^n. We illustrate the definition by showing that the unit sphere

$$S^n = \{x \in \mathbb{R}^{n+1} \ : \ |x| = 1\}$$

is a C^∞ n-dimensional submanifold of $\mathbb{R}^{n+1}$, where $|x| := (x_1^2 + \cdots + x_{n+1}^2)^{1/2}$. In fact, we will show that $2(n+1)$ parametrizations are enough to cover the whole sphere. For each $i = 1, \ldots, n+1$, we set

$$H_i^+ = \{y \in \mathbb{R}^{n+1} : x_i > 0\}$$
$$H_i^- = \{y \in \mathbb{R}^{n+1} : x_i < 0\}.$$

These are open half-spaces having boundary $x_i = 0$. They give rise to open the sets $U_i^+ = H_i^+ \cap S^n$ and $U_i^- = H_i^+ \cap S^n$ in S^n such that

$$\bigcup_{i=1}^{n+1} (U_i^+ \cup U_i^-) = S^n.$$

(This is clear since each point of S^n must have some coordinate different from zero, say x_i, which is either positive or negative, so that the point lies either in U_i^+ or U_i^-.)

Each $U_i^\pm$ is equipped with a C^∞ parametrization $\varphi_i^\pm : B \to U_i^\pm$ defined by:

$$(x_1, \ldots, x_n) \mapsto (x_1, \ldots, x_{i-1}, \pm\sqrt{1 - |x|^2}, x_i, \ldots, x_n).$$

Here $B = \{x \in \mathbb{R}^n : |x| < 1\}$ denotes the open ball.

Exercise 3.4.1. Verify that the maps $\varphi_i^\pm$ are indeed parametrizations, i.e., they are immersions over B and that $\varphi_i^\pm$ is a homeomorphism from B onto $U_i^\pm$.

We now discuss changes of parametrization. Let M be an m-dimensional submanifold of $\mathbb{R}^n$ of class C^k, $k \geq 1$. Let U be an open subset of M (in the relative topology obtained by the inclusion $M \subset \mathbb{R}^n$) and $\varphi : U_0 \to U$ a parametrization of U. Via φ each point of U becomes specified by m coordinates:

$$(x_1, \dots, x_m) \in U_0 \mapsto p = \varphi(x_1, \dots, x_m) \in U.$$

If $g : V_0 \to U_0$ is a C^k diffeomorphism from another open set $V_0 \subset \mathbb{R}^m$ onto U_0, then $\varphi \circ g : V_0 \to U$ is also a parametrization of U. The function g is called a *change of coordinates*.

Change of coordinates is the only way to obtain new parametrizations. To see this, suppose that $\varphi : U_0 \to U$ and $\psi : V_0 \to V$ are parametrizations of open subsets U and V of M such that $U \cap V$ is nonempty. The map

$$g := \psi^{-1} \circ \varphi : \varphi^{-1}(U \cap V) \to \psi^{-1}(U \cap V)$$

is clearly a homeomorphism between open sets in $\mathbb{R}^m$. Differentiability of g is not immediate since ψ^{-1} is not defined on an open subset of $\mathbb{R}^n$. However, the next proposition takes care of this difficulty.

Proposition 3.4.2. Let V_0 be an open subset of $\mathbb{R}^m$ and $\psi : V_0 \to V$ a C^k parametrization of $V \subset \mathbb{R}^n$. Then, given an open $U_0 \subset \mathbb{R}^r$ and $f : U_0 \to V$ of class C^k, the composition $\psi^{-1} \circ f : U_0 \to V_0$ is C^k and for each $x \in U_0$ and $z = (\psi^{-1} \circ f)(x)$ we have $d(\psi^{-1} \circ f)_x = (d\psi_z)^{-1} \circ df_x$.

Proof. Once we know that $h = \psi^{-1} \circ f$ is differentiable we can apply the chain rule to $\psi \circ h = f$ to obtain the last claim.

We now prove that h is C^k. Since ψ is an (injective) C^k immersion, for each $p \in V$ there exist an open subset $Z \subset \mathbb{R}^n$ containing p and a C^k map $g : Z \to \mathbb{R}^m$ such that $g|_{V \cap Z} = \psi^{-1}$. (This is an immediate consequence of the theorem on local form of immersions. The map g corresponds to the composition of the diffeomorphism h obtained in that theorem with the projection $(x, y) \mapsto x$.)

Let p be an arbitrary point of $f(U_0) \subset V$. Then $\psi^{-1} \circ f = g \circ f : f^{-1}(f(U_0) \cap Z) \subset \mathbb{R}^r \to \mathbb{R}^m$. Consequently, $\psi^{-1} \circ f$ is C^k, since both f and g are. $\square$

Corollary 3.4.3. Let U_0 and V_0 be open subsets of $\mathbb{R}^m$ and $\varphi : U_0 \to V$ and $\psi : V_0 \to V$ two C^k parametrizations of the same set $V \subset V$. The the change of coordinates $g = \psi^{-1} \circ \varphi$ is a C^k diffeomorphism.

Up until now the notion of a differentiable map was only defined for functions whose domain is an open set in Euclidean space. The above corollary allows one to extend the notion to maps defined on manifolds, as follows. Let M be a C^k m-dimensional submanifold of $\mathbb{R}^n$. We say that $f : M \to \mathbb{R}^s$ is differentiable at $p \in M$ if there exists a parametrization $\varphi : U_0 \to U$ of class C^k, $k \geq 1$, with $p \in U$, such that $f \circ \varphi : U_0 \to \mathbb{R}^s$ is differentiable at $p_0 = \varphi^{-1}(p) \in U_0$. The proposition shows that this definition does not depend on the choice of parametrization since, given any other parametrization ψ (of another neighborhood of p) we have that $f \circ \psi = (f \circ \varphi) \circ (\varphi^{-1} \circ \psi)$ is differentiable at $\varphi^{-1}(p)$.

The notion of C^k differentiability also extends without difficulty to manifold, so long as M itself is C^k.

If $M \subset \mathbb{R}^r$ and $N \subset \mathbb{R}^s$ are C^k manifolds of dimensions m and n respectively, we say that $f : M \to N$ is differentiable at $p \in M$ when considered as a map from M into $\mathbb{R}^s$ f is differentiable at p. The definition of a C^k map f from M into N is analogous.

The next proposition shows that differentiability of $f : M \to N$ can also be defined purely in terms of parametrizations into M and N, without explicit reference to the ambient Euclidean spaces.

PROPOSITION 3.4.4. A map $f : M \to N$ is of class C^k if and only if for all $p \in M$ there are neighborhoods $U \subset M$ of p and $V \subset N$ of $f(p)$ satisfying $f(U) \subset V$, and C^k parametrizations $\varphi : U_0 \to U$ and $\psi : V_0 \to V$, such that

$$\psi^{-1} \circ f \circ \varphi : U_0 \to V_0$$

is C^k.

PROOF. Let $f : M \to N$ be of class C^k. Given $p \in M$, let $\psi : V_0 \to V \subset N$ be a C^k parametrization with $f(p) \in V$, $V_0 \subset \mathbb{R}^n$. Since f is continuous there is a parametrization $\varphi : U_0 \to U \subset$ with $p \in U$ such that $f(U) \subset V$. Since f is C^k, the map $f \circ \varphi : U_0 \mapsto V$ is (by definition) C^k. By the previous proposition it follows that $\psi^{-1} \circ f \circ \varphi : U_0 \to V_0$ is C^k. The converse is easier and is left to the reader. $\square$

COROLLARY 3.4.5. If $f : M \to N$ and $g : N \to P$ are C^k maps, then the composition $g \circ f : M \to P$ is also C^k.

CHAPTER 4

Differentiable Manifolds

DEFINITION 4.0.6 (Differentiable Structure). An n-dimentional *differentiable manifold* is a set M together with a family of injective maps $f_\alpha : U_\alpha \subset \mathbb{R}^n \to M$ from open sets U_α in $\mathbb{R}^n$ into M such that:

(1) the union of all $f_\alpha(U_\alpha)$ is M;
(2) for each pair α, β for which the intersection $W = f_\alpha(U_\alpha) \cap f_\beta(U_\beta)$ is non-empty, the sets $f_\alpha^{-1}(W)$ and $f_\beta^{-1}(W)$ are open sets in $\mathbb{R}^n$ and the maps $f_\alpha^{-1} \circ f_\beta$ and $f_\beta^{-1} \circ f_\alpha$ are differentiable;
(3) the family $\{(U_\alpha, f_\alpha)\}$ is maximal relative to (1) and (2).

A family $\{(U_\alpha, f_\alpha)\}$ satisfying properties (1) and (2) is called a *differentiable structure* on M. The pair (U_α, f_α), with $p \in f_\alpha(U_\alpha)$ is called a *parametrization* (or a *coordinate system*) of M at p, and $f_\alpha(U_\alpha)$ is called a *coordinate neighborhood* of p

Condition (3) is a useful technical assumption. A differentiable structure on M can always be extended so as to satisfy (3). This is achieved by adjoining all parametrizations for which (2) still holds for the extended differentiable structure.

[We also assume in these notes that M (with the topology generated by open sets of the form $f_\alpha^{-1}(W)$, where W is open in $\mathbb{R}^n$) is a Hausdorff space and that the topology of M admits a countable basis.]

We will see in a number of examples that manifolds often arise as the "space of configurations" of some physical or mathematical object.

DEFINITION 4.0.7. Let M and N be differentiable manifolds of dimensions m and n, respectively. Then a map $\varphi : M \to N$ is aid to be differentiable at a point $p \in M$ if there exists a parametrization (U, f) on M with $p \in f(U)$, and a parametrization (V, g) on N with $\varphi(f(U)) \subset g(V)$ such that

$$g^{-1} \circ \varphi \circ f : U \subset \mathbb{R}^n \to \mathbb{R}^m$$

is differentiable at $f(p)$. If φ is differentiable on all points of an open set $W \subset M$ we say that φ is differentiable on W. If φ is invertible and both it and its inverse are everywhere differentiable, we say that φ is a diffeomorphism between M and N.

The previous definition makes sense for any order of differentiation. The set of all real valued functions on M that are differentiable, with continuous derivatives, up to order k, will be denoted by $C^k(M)$.

b. A 2-torus in $\mathbb{R}^3$. The image of the function $f : \mathbb{R}^2 \to \mathbb{R}^3$ such that

$$f(u, v) = ((2 + \cos v) \cos u, (2 + \cos v) \cos u, \sin v)$$

is the surface of a torus in $\mathbb{R}^3$. A differentiable structure on $T = f(\mathbb{R}^2)$ is obtained by taking $\{(U_i, f_i), i = 1, \dots, 4\}$, where f_i is the restriction of f to U_i and

$$\begin{aligned} U_1 &= \{(u,v) : -\pi < u < \pi, \quad -\pi < v < \pi\} \\ U_2 &= \{(u,v) : -\pi < u < \pi, \quad 0 < v < 2\pi\} \\ U_3 &= \{(u,v) : \quad 0 < u < 2\pi, \ -\pi < v < \pi\} \\ U_4 &= \{(u,v) : \quad 0 < u < 2\pi, \quad 0 < v < 2\pi\}. \end{aligned}$$

EXERCISE 4.0.8. Show that $\{(U_i, f_i)\}$ is a differentiable structure on the torus. What are the change of coordinate functions $f_i^{-1} \circ f_j$?

If M and N are diffeomorphic manifolds (that is, if there is a diffeomorphism between the two), then M and N are indistinguishable as far as their abstract manifold structure is concerned. For example, the 2-torus could be given the following characterization. Regard $\mathbb{Z}^2$ as a group of integer translations on the vector space $\mathbb{R}^2$. Denote by $x + \mathbb{Z}^2$ the *orbit* of $x = (x, y)$ under the *action* of $\mathbb{Z}^2$ on $\mathbb{Z}^2$. In other words, $x + \mathbb{Z}^2$ is the set of all $x + n$ where $n = (n_1, n_2)$ ranges over $\mathbb{Z}^2$. Let

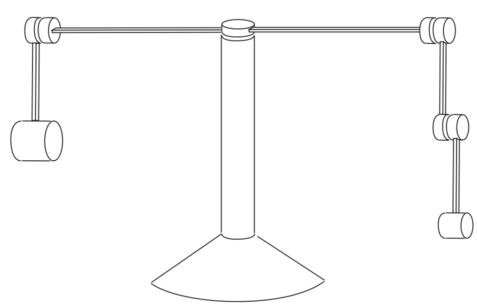
$$M = \{x + \mathbb{Z}^2 : x \in \mathbb{R}^2\}$$

denote the set of all orbits.

EXERCISE 4.0.9. Show that M admits a unique smooth manifold structure for which the projection $x \in \mathbb{R}^2 \mapsto x + \mathbb{Z}^2 \in M$ is a differentiable map. Also show that M and T are diffeomorphic.

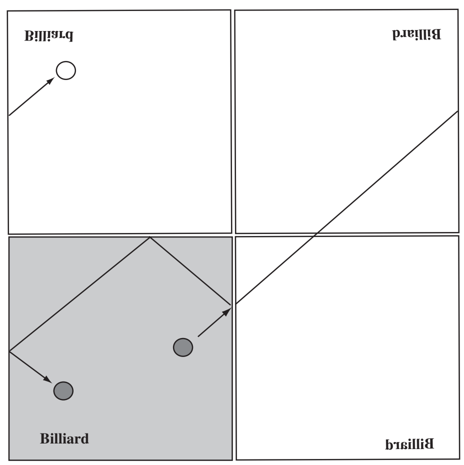
It should be clear that $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$ makes sense and is a manifold for all $n \geq 1$.

The mechanical system represented in the following picture consists of a double pendulum and a simple pendulum suspended on the opposite sides of a rod that can rotate freely around a vertical pole. The description of a configuration of the system involves 4 parameters (4 "degrees of freedom"). The set of configurations is globally parametrized by $\mathbb{T}^4$.



EXERCISE 4.0.10 (Product Manifolds). Let M and N be differentiable manifolds and let $\{U_i, f_i\}$, $\{V_j, g_j\}$ be differentiable structures on M and N, respectively. Consider the cartesian product $M \times N$ and the mappings $f_{ij}(p, q) = (f_i(p), g_j(q))$, $p \in U_i, q \in V_j$. Show that $\{U_i \times V_j, f_{ij}\}$, is a differentiable structure on $M \times N$ in which the projections $\pi_1 : M \times N \to M$ and $\pi_2 : M \times N \to N$ are differentiable. With this differentiable structure $M \times N$ is called the *product manifold of* M and N.

Another, very different, situation for which the space of configurations can also be described by a torus is the *ball on a billiard table* system described in the next figure. A solid ball moves on a rectangular billiard table along straight lines betwee collisions with a wall. After hitting a wall that ball bounces off elastically according to the usual equal angles rule.



Although the space of configurations of the system is a rectangle (a manifold with boundary), it may be more convenient to represent the motion on a slightly modified space where all the trajectories of the ball are rectilinear. This is done by first "unfolding" the rectangle by reflecting the table on its boundary segments (thus rescalling the rectangle by a factor of 2), and then gluing the opposite sides so as to form the torus $\mathbb{T}^2$. This is indicated in the next figure. The torus thus obtained can be regarded as $\mathbb{R}^2/\mathbb{Z}^2$.

c. The real projective plane. Let $M = \mathbb{R}P^2$ denote the *real projective plane*, that is, the set of all lines through the origin in $\mathbb{R}^3$. We show that $\mathbb{R}P^2$ is a two-dimensional manifold.

Let $\pi : \mathbb{R}^3 - \{(0,0,0)\} \to \mathbb{R}P^2$ be the map that associates to each nonzero $x = (x, y, z) \in \mathbb{R}^3$ the line $\pi(x) = \{tx : t \in \mathbb{R}\}$. Define

$$V_1 = \{p \in \mathbb{R}P^2 : p = \pi(x, y, z), x \neq 0\}$$
$$V_2 = \{p \in \mathbb{R}P^2 : p = \pi(x, y, z), y \neq 0\}$$
$$V_3 = \{p \in \mathbb{R}P^2 : p = \pi(x, y, z), z \neq 0\}$$

and maps $f_i : \mathbb{R}^2 \to V_i$ by

$$f_1(u, v) = \pi(1, u, v)$$
$$f_2(u, v) = \pi(u, 1, v)$$
$$f_3(u, v) = \pi(u, v, 1).$$

Notice that $f_1(u, v) = \pi(1, u, v) = \pi(\frac{1}{u}, 1, \frac{v}{u}) = f_2(\frac{1}{u}, \frac{v}{u})$, so that

$$f_2^{-1} \circ f_1(u, v) = (\frac{1}{u}, \frac{v}{u}).$$

This is a differentiable map since $u \neq 0$ for $(u, v) \in f_1^{-1}(V_1 \cap V_2)$.

Exercise 4.0.11. Let $U_i = \mathbb{R}^2$. Show that $\{(U_1, f_1), (U_2, f_2), (U_3, f_3)\}$ is a differentiable structure on $\mathbb{R}P^2$.

Exercise 4.0.12. Let $F : \mathbb{R}^3 \to \mathbb{R}^4$ be given by

$$F(x, y, z) = (x^2 - y^2, xy, xz, yz).$$

Let S^2 be the unit sphere in $\mathbb{R}^3$ with center at the origin. Use F to obtain an embedding of $\mathbb{R}P^2$ into $\mathbb{R}^4$.

Exercise 4.0.13. Show that for any point p in the projective plane, the open set $\mathbb{R}P^2 - \{p\}$ is diffeomorphic with the Möbius band.

d. Submanifolds of $\mathbb{R}^n$. Let M be a C^k, $k \geq 1$, submanifold of $\mathbb{R}^n$ of dimension m. Let p be a point in M and $\varphi : U_0 \to U$ a parametrization of a neighborhood U of p. Suppose that $p = \varphi(x)$. The *tangent space* to M at p is the m-dimensional vector space

$$TM_p := d\varphi_x \mathbb{R}^m.$$

The vectors $d\varphi_x e_i$, $i = 1, \ldots, m$, form a basis for TM_p.

Although we have used a parametrization of a neighborhood of p to define TM_p, the resulting vector space is actually independent of the choice made. This can be seen as follows. Let $\psi : V_0 \to V$ be another parametrization at p. Let

$$g = \psi^{-1} \circ \varphi : \varphi^{-1}(U \cap V) \to \psi^{-1}(U \cap V)$$

be the coordinate change, with $p = \varphi(x) = \psi(z)$. Since g is a diffeomorphisms, we must have $\mathbb{R}^m = df_x \mathbb{R}^m$. Therefore, by the chain rule we have

$$d\varphi_x \mathbb{R}^m = d\psi_x dg_x \mathbb{R}^m = d\psi_z \mathbb{R}^m.$$

Proposition 4.0.14. The elements of TM_p are the velocity vectors at p of differentiable paths in M that pass through p. More precisely, TM_p is the set of all $v \in \mathbb{R}^n$ such that there exists a differentiable path $c : (-\epsilon, \epsilon) \to M$ with $p = c(0)$ and $v = c'(0)$.

PROOF. Let $v \in TM_p$. By definition, there exist a parametrization $\varphi : U_0 \to U$ with $\varphi(x) = p$ and a vector $u \in \mathbb{R}^m$ such that

$$v = d\varphi_x u = \lim_{t\to 0} \frac{\varphi(x+tu) - \varphi(x)}{t}.$$

Choose ϵ sufficiently small such that the straight line $x+tu$, $-\epsilon < t < \epsilon$, is contained in U_0. Therefore v is the velocity vector at $t = 0$ of the path $c(t) = \varphi(x+tu)$.

For the converse, let $c : (-\epsilon, \epsilon) \to M$ be a differentiable path with $c(0) = p$. Let $\varphi : U_0 \to U$ be any parametrization with $p \in U$. By taking ϵ sufficiently small we may assume without loss of generality that $c(t) \in U$ for all $t \in (-\epsilon, \epsilon)$. It follows that the path $\varphi^{-1} \circ c : (-\epsilon, \epsilon) \to U_0$ is differentiable and, setting $u := (\varphi^{-1} \circ c)'(0) = [d\varphi_x]^{-1} c'(0)$, we have $c'(0) = d\varphi_x u$. □

The vector space TM_p is a vector subspace of $\mathbb{R}^n$, hence it passes through the origin. It is however helpful to imagine the tangent space translated to p, i.e., to consider the affine subspace $p + TM_p$.

Let M and N be differentiable submanifolds of $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively, and let $f : M \to N$ be differentiable at a point $p \in M$. The *differential* of f at p is the linear transformation

$$df_p : TM_p \to TN_{f(p)}$$

that associates to each $v = \gamma'(0) \in TM_p$ the vector $df_p v = (f \circ \gamma)'(0) \in TN_{f(p)}$, where $\gamma(t)$ is a differentiable path such that $\gamma(0) = p$ and $\gamma'(0) = v$.

We need to check that df_p is indeed a linear transformation. It will be clear from the discussion that df_p reduces to the already defined differential of functions from open subsets of Euclidean space. Let $\varphi : U_0 \subset \mathbb{R}^m \to M$ be a differentiable parametrization of a neighborhood U of p and let $\psi : V_0 \to V \subset N$ be a parametrization of a neighborhood V of $f(p)$ in N. By reducing the size of U we may assume that $f(U) \subset V$, so that we can form the composition $g := \psi^{-1} \circ \varphi : U_0 \to V_0$. Then

$$\begin{aligned}(\psi^{-1} \circ f \circ \gamma)'(0) &= (\psi^{-1} \circ f \circ \varphi \circ \varphi^{-1} \circ \gamma)'(0) \\ &= d(\psi^{-1} \circ f \circ \varphi)_{\varphi^{-1}(\gamma(0))} (\varphi^{-1} \circ \gamma)'(0).\end{aligned}$$

This shows that $d\psi^{-1}_{f(p)} df_p v = dg_{\varphi^{-1}(p)} d\varphi_p^{-1} v$, which only depends on the velocity vector of γ at p. Moreover

$$df_p v = d\psi_{\psi^{-1}(f(p))} dg_{\varphi^{-1}(p)} d\varphi_p^{-1} v$$

therefore df_p is a linear transformation.

EXERCISE 4.0.15. Show that the chain rule still holds. More precisely, if M, N, P are manifolds, $f : M \to N$ is a differentiable map at $p \in M$ and $g : N \to P$ is a differentiable map at $f(p) \in N$, then $g \circ f : M \to P$ is differentiable at p and

$$d(g \circ f)_p = dg_{f(p)} \circ df_p : TM_p \to TP_{g(f(p))}.$$

EXERCISE 4.0.16. Let A be a symmetric n by n real matrix and define

$$f(x) = \langle Ax, x\rangle,$$

for $x \in \mathbb{R}^n$. By restriction to the unit sphere, we obtain a differentiable function $f : S^{n-1} \to \mathbb{R}$.

(1) Show that the differential of f at a point $p \in S^{n-1}$ is given by

$$df_p v = \langle Ap, v\rangle.$$

restriction of A to TS^{n-1}_p.

(2) If $p \in S^{n-1}$ is a point where f attains its maximum value, show that $df_p v = 0$ for all $v \in TS_p^{n-1}$.
(3) Show that at a critical point p of f (i.e., a point where $df_p = 0$, such as a point of maximum), $Ap = \lambda p$ for some real number λ. Therefore, each critical point of f is an eigenvector of A.
(4) Show that the orthogonal subspace to p is sent to itself under the linear transformation A. Use this fact and induction to show that any symmetric n by n real matrix admits an orthonormal basis of eigenvectors.

e. Graphs and level sets. Let U be an open subset of $\mathbb{R}^m$ and $f : U \to \mathbb{R}^n$ a C^k map. The graph of f

$$G(f) = \{(x, f(x)) : x \in U\}$$

is a C^k m-dimensional submanifold of $\mathbb{R}^{m+n}$. In fact, $\varphi : U \to G(f)$ defined by $\varphi(x) = (x, f(x))$, is a parametrization of the whole $G(f)$.

Although it is not the case that all submanifolds of $\mathbb{R}^n$ are graphs, this is true locally. More precisely, the following proposition holds.

PROPOSITION 4.0.17. Let M be a C^k m-dimensional submanifold of $\mathbb{R}^n$. Then every $p \in M$ has a neighborhood V parametrized by a C^k map $\psi : V_0 \to V$ of the form $\psi(x) = (x, f(x))$ for $x \in V_0 \subset \mathbb{R}^m$.

PROOF. Let $\varphi : U_0 \subset \mathbb{R}^m \to U \subset M$ a parametrization of a neighborhood U of $p = \varphi(x)$. We fix a direct sum decomposition $\mathbb{R}^n = \mathbb{R}^m \oplus \mathbb{R}^{n-m}$ having the property that the projection $\pi : \mathbb{R}^n \to \mathbb{R}^m$ maps TM_p isomorphically onto $\mathbb{R}^m$.

Define $\eta := \pi \circ \varphi : U_0 \subset \mathbb{R}^m \to \mathbb{R}^m$. Then $d\eta_x = \pi \circ d\varphi_x$ is an isomorphism, so by the inverse function theorem η is a C^k diffeomorphism from a possibly smaller neighborhood U_1 of x onto a neighborhood V_0 of $\pi(x)$. Denote by $\xi = \eta^{-1} : V_0 \to U_1$ the inverse diffeomorphism. Then

$$\psi = \varphi \circ \xi : V_0 \subset \mathbb{R}^m \to V = \psi(V_0) \subset \mathbb{R}^n$$

is a new parametrization of a neighborhood of p. It follows from the relation $\pi \circ \psi = \pi \circ (\varphi \circ \xi) = (\pi \circ \varphi) \circ \xi = \eta \circ \xi = \mathrm{id}_{V_0}$ that the first coordinate of $\psi(x)$ relative to the decomposition $\mathbb{R}^n = \mathbb{R}^m \oplus \mathbb{R}^{n-m}$ is x. Let $f(x)$ denote the second coordinate. Then $\psi(x) = (x, f(x))$, $x \in V_0$. Note that $\psi = (\pi|_V)^{-1} : V_0 \to V$, that is, the parametrization that makes V a graph is simply the local inverse of the projection $\pi : \mathbb{R}^m \oplus \mathbb{R}^{n-m} \to \mathbb{R}^m$ that takes TM_p onto $\mathbb{R}^m$ isomorphically. □

Consider the C^∞ function $f(x_1, \ldots, x_{n+1}) = x_1^2 + \cdots + x_{n+1}^2$. The set defined implicitly by the equation $f(x_1, \ldots, x_{n+1}) = 1$ is the unit sphere S^n. Similarly, the function

$$g(x_1, \ldots, x_{n+1}) = x_1^2 + \cdots + x_n^2 - x_{n+1}^2$$

defines implicitly an n-dimensional hypersurface by $g(x_1, \ldots, x_{n+1}) = 1$ which, for $n = 2$, is a hyperboloid of two sheets. The equation $g(x_1, x_2, x_3) = 0$ defines a pair of cones touching at the origin, hence cannot be a manifold. (No neighborhood of the origin admits a parametrization.) The next proposition gives sufficient conditions for an equation $f(x) = c$ to define a manifold.

PROPOSITION 4.0.18. Let $U \subset \mathbb{R}^{m+n}$ be an open set and $f : U \to \mathbb{R}^n$ a C^k map. Fix $c \in \mathbb{R}^n$ and consider the set

$$M = \{p \in U : f(p) = c \text{ and } df_p : \mathbb{R}^{m+n} \to \mathbb{R}^n \text{ is surjective}\}.$$

Then

(1) M is open in $f^{-1}(c)$
(2) Assuming that M is nonempty, M is a C^k m-dimensional submanifold of $\mathbb{R}^{m+n}$.
(3) $TM_p = \ker(df_p)$ for all $p \in M$.

Proof. The verification of the first claim is immediate.

Let $p \in M$. By the implicit function theorem we can find a decomposition $\mathbb{R}^{m+n} = \mathbb{R}^m \oplus \mathbb{R}^n$ with $p = (x_0, y_0)$, neighborhoods $Z \subset \mathbb{R}^{m+n}$ of p, $V \subset \mathbb{R}^m$ of x_0, and a C^k map $\xi : V \to \mathbb{R}^n$ such that the graph of ξ satisfies $G(\xi) = Z \cap f^{-1}(c)$. Therefore $\varphi : V \to Z \cap f^(c)$, given by $\varphi(x) = (x, \xi(x))$ is a C^k parametrization of an open neighborhood of p in $f^{-1}(c)$. Moreover df_q is surjective for each $q \in Z$ so that $Z \cap f^{-1}(c) \subset M$. This concludes the proof of claim 2.

Let $v \in TM_p$. Consider a path $\gamma : (-\epsilon, \epsilon) \to M$ such that $\gamma(0) = p$ and $\gamma'(0) = v$. Then $df_p v = df_{\gamma(0)}\gamma'(0) = (f \circ \gamma)'(0) = 0$ since $f \circ \gamma$ is constant. Therefore, v lies in the kernel of df_p. Since TM_p and $\ker(df_p)$ are m-dimensional subspaces of $\mathbb{R}^{m+n}$ and $TM_p \subset \ker(df_p)$ we conclude that $TM_p = \ker(df_p)$. □

Let $f : U \to \mathbb{R}^n$ be a differentiable map from an open $U \subset \mathbb{R}^m$. A point $c \in \mathbb{R}^n$ is called a *regular value* of f if, for each $x \in U$ such that $f(x) = c$ the differential $df_x : \mathbb{R}^m \to \mathbb{R}^n$ is surjective. If there are no $x \in U$ for which $f(x) = c$ then c is trivially a regular value. When $n = 1$, the linear functional $df{+}p : \mathbb{R}^m \to \mathbb{R}$ is either 0 or surjective. In this case, the real number c is a regular value if and only if $df_p \neq 0$ for all $x \in f^{-1}(c)$. For example, consider $g(x_1, \dots, x_n, x_{n+1}) = x_1^2 + \cdots + x_n^2 - x_{n+1}^2$. Then, representing by $\{dx_1, \dots, dx_{n+1}\}$ the basis of $(\mathbb{R}^{n+1})^*$ dual to the canonical basis of $\mathbb{R}^{n+1}$, we have

$$df_x = 2x_1 dx_1 + \cdots + 2x_n dx_n - 2x_{n+1} dx_{n+1}.$$

It follows that $df_x = 0$ if and only if $x_1 = \cdots = x_n = x_{n+1} = 0$. Since $f(0, \dots, 0) = 0$ we conclude that 0 is the only nonregular value of f.

Using the definition of a regular value, the previous proposition can be restated as follows.

Theorem 4.0.19. Let $U \subset \mathbb{R}^n$ be an open set and $f : U \to \mathbb{R}^{n-m}$ a C^k map, $k \geq 1$. if $c \in \mathbb{R}^{n-m}$ is a regular value of f, either f^{-1} is empty or it is a C^k m-dimensional submanifold of $\mathbb{R}^n$. Moreover, for each $p \in f^{-1}(c)$, the tangent space $T[f^{-1}(c)]_p$ is the kernel of the differential $df_p : \mathbb{R}^n \to \mathbb{R}^{n-m}$.

Exercise 4.0.20. Show that the unit sphere is an n-dimensional submanifold of $\mathbb{R}^{n+1}$ by showing that 1 is a regular value of $f(x) = |x|^2$, $x \in \mathbb{R}^{n+1}$. Verify by a direct computation that TS^n_p is the kernel of df_p.

Exercise 4.0.21. Let U be the complement of the z-axis in $\mathbb{R}^3$ and define

$$f(x, y, z) = z^2 + (\sqrt{x^2 + y^2} - 2)^2.$$

Show that df_p is nonzero for all p outside the circle

$$S = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 4, z = 0\}$$

and that 0 is the only nonregular value of f. (For $0 < c < 4$, $f^{-1}(c)$ is the torus generated by rotating the circle of radius $\sqrt{c}$ whose center lies on S.)

f. Matrix groups. A subset $G \subset M(n, \mathbb{R})$ that is a group under matrix multiplication will be called a (real) *linear group*. We will also refer to it as a (real) *matrix group*.

We say that G is a (linear) *Lie group* if it is also a smooth submanifold of $M(n, \mathbb{R})$. In general, a Lie group is a group which is also a differentiable manifold such that the group operations of multiplication and inverse are differentiable.

1. *The General Linear Group.* $GL(n, \mathbb{R})$ is the group of all invertible $n \times n$ real matrices. It is the subset of $M(n, \mathbb{R})$ consisting of A having positive determinant. Since the determinant is a continuous function on $M(n, \mathbb{R})$, $GL(n, \mathbb{R})$ is an open subset, hence a manifold. The operations of matrix multiplication and matrix inverse are differentiable (these are rational functions on the matrix entries). Therefore $GL(n, \mathbb{R})$ is a Lie group.

2. *The Special Linear Group.* Let $SL(n, \mathbb{R})$ be the subgroup of $GL(n, \mathbb{R})$ consisting of matrices of determinant 1.

We first show that $SL(n, \mathbb{R})$ is a smooth submanifold of $M(n, \mathbb{R})$. Consider the function $D : M(n, \mathbb{R}) \to \mathbb{R}$ that associates to each n by n matrix its determinant. Then $SL(n, \mathbb{R}) = D^{-1}(1)$. If we show that 1 is a regular value of D, the implicit function theorem can be used to conclude that $D^{-1}(1)$ is a smooth manifold of dimension $n^2 - 1$. To see that 1 is a regular value of D, is suffices to check that dD_A is not zero for all $A \in SL(n, \mathbb{R})$. The next exercise will be useful.

EXERCISE 4.0.22. Show that $dD_A X = D(A)\mathrm{Tr}(XA^{-1})$.

Choosing $X = A$, this expression yields $dD_A A = n \neq 0$.

3. *The Orthogonal Group.* $O(n)$ is the group of all matrices A such that $AA^t = I$. We will show that it is a differentiable submanifold of $M(n, \mathbb{R}) = \mathbb{R}^{n^2}$ of dimension $n(n-1)/2$. Let $S_n(\mathbb{R})$ denote the vector space, isomorphic to $\mathbb{R}^{n(n+1)/2}$, of symmetric matrices; i.e, $S_n(\mathbb{R})$ consists of matrices $A \in M(n, \mathbb{R})$ such that $A^t = A$, where A^t denotes the transpose of A. Define a C^∞ map

$$f : M(n, \mathbb{R}) \to S_n(\mathbb{R}), \quad f(A) = AA^t.$$

To show that $O(n)$ is a smooth manifold, it suffices to show that $I \in S_n(\mathbb{R})$ is a regular value of f. It will then follow that the dimension is $n^2 - n(n+1)/2 = n(n-1)/2$. The following exercises will be useful for proving that I is indeed a regular value.

EXERCISE 4.0.23. Show that any $A \in GL(n, \mathbb{R})$ gives rise to a diffeomorphism

$$L_A : M(n, \mathbb{R}) \to M(n, \mathbb{R})$$

defined by $L_A(B) = AB$. Show the same for

$$R_A : M(n, \mathbb{R}) \to M(n, \mathbb{R})$$

defined by $R_A(B) = BA$.

The diffeomorphism L_A and R_A introduced in the previous exercise are called *left-multiplication* and *right-multiplication* by A, respectively.

EXERCISE 4.0.24. If $A \in O(n)$, then L_A maps any neighborhood U of the identity matrix $I \in M(n, \mathbb{R})$ diffeomorphically onto a neighborhood of A so that $O(n) \cap U$ is sent to $O(n) \cap L_A(U)$. (A similar statement holds for R_A.)

EXERCISE 4.0.25. If $f(A) = A^t A$ for $A \in M(n, \mathbb{R})$, show that $df_I V = V^t + V$, where I is the identity matrix.

We can now show that $O(n)$ is a C^∞ manifold as follows. By the second of the last three exercises it suffices to show that $O(n) \cap U$ is a submanifold of U for some neighborhood U of the identity matrix I. Since by the last exercise $df_I V = V + V^t$, we see that $df_I : M(n, \mathbb{R}) \to S_n(\mathbb{R})$ is surjective since for any $W \in S_n(\mathbb{R})$ we have $df_I W/2 = W$. Therefore, we can apply the implicit function theorem.

PROPOSITION 4.0.26. $O(n)$ and $SO(n)$ are compact Lie groups.

PROOF. We already know that $O(n)$ is a smooth submanifold of $M(n, \mathbb{R})$. That is is compact is easily seen, as follows. The euclidian norm on $M(n, \mathbb{R})$ can be written as

$$\|A\| = \mathrm{Tr}(AA^t).$$

Therefore the orthogonal group is bounded (contained in a sphere of radius $\sqrt{n}$ in $\mathbb{R}^{n^2}$) and closed, hence compact.

$SO(n)$ is compact, being closed in $O(n)$, and it is also open in $O(n)$ since it is $O(n) \cap D^{-1}(0, \infty)$ and the determinant D is a continuous function. Since an open subset of a differentiable manifold is also a differentiable manifold, $SO(n)$ is a compact manifold. □

There are local parametrizations of $SO(3)$ well-known from elementary mechanics courses, such as the Euler angles. A particularly convenient set of parametrizations for the purpose of obtaining a manifold structure is the following, known as *Cayley parametrizations*. To each $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ associate the skew-symmetric matrix

$$A(x) = \begin{pmatrix} 0 & x_1 & x_2 \\ -x_1 & 0 & x_3 \\ -x_2 & -x_3 & 0 \end{pmatrix}.$$

On a small enough neighborhood U of the origin, we can assume that $A(x) - I$ as well as $I + (A(x) - I)^{-1}(A(x) + I)$ are non-singular. Define $f(A) = (I - A)^{-1}(I + A)$ and, for each $R \in SO(3)$, write $f_R(A) = Rf(A)$.

EXERCISE 4.0.27. Show that $\{(U, f_R) : R \in SO(3)\}$ is a differentiable structure on $SO(3)$ and that the action of $SO(3)$ on $\mathbb{R}^3$ is differentiable with respect to it.

4. *The Group of Euclidian Isometries.* A map $f : \mathbb{R}^n \to \mathbb{R}^n$ is an *isometry* if for any $p, q \in \mathbb{R}^n$, $\|f(p) - f(q)\| = \|p - q\|$ (with the ordinary root-of-sum-of-squares norm). We will see in a moment that any isometry of $\mathbb{R}^n$ has the form $f(p) = Ap + q$ where $A \in O(n)$. The set of all isometries of $\mathbb{R}^n$ will be denoted $E(n)$.

EXERCISE 4.0.28. Show that $E(n)$ is a group under composition. Show that it is isomorphic to the semidirect product $O(n) \ltimes \mathbb{R}^n$. (The latter is defined as follows. As a set, it is the product $O(n) \times \mathbb{R}^n = \{(A, u) : A \in O(n), u \in \mathbb{R}^n\}$, with multiplication given by

$$(A_2, u_2)(A_1, u_1) = (A_2 A_1, A_2 u_1 + u_2).$$

The group $E(n)$ can also be written as a subgroup of $GL(n+1, \mathbb{R})$ by means of the correspondence:

$$(A, u) \in O(n) \ltimes \mathbb{R}^n \mapsto \begin{pmatrix} A & u \\ 0 & 1 \end{pmatrix} \in GL(n+1, \mathbb{R}).$$

Note that

$$\begin{pmatrix} A_2 & u_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A_1 & u_1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} A_2 A_1 & A_2 u_1 + u_2 \\ 0 & 1 \end{pmatrix}$$

Therefore, the correspondence is a group homomorphism (with trivial kernel) and realizes $O(n) \ltimes \mathbb{R}^n$ as a subgroup of $GL(n+1, \mathbb{R})$. Also note that $SO(n) \ltimes \mathbb{R}^n$ defines a subgroup of $E(n)$, which we denote $SE(n)$. Both $E(n)$ and $SE(n)$ can be realized as smooth submanifolds of $M(n+1, \mathbb{R})$.

If we identify $\mathbb{R}^n$ with the affine subspace of $\mathbb{R}^{n+1}$ consisting of points whose $n+1$st coordinate is 1, then the transformation

$$p \mapsto Ap + u$$

of $\mathbb{R}^n$ determined by $(A, u) \in E(n)$ is expressed as

$$\begin{pmatrix} A & u \\ 0 & 1 \end{pmatrix} \begin{pmatrix} p \\ 1 \end{pmatrix} = \begin{pmatrix} Ap + u \\ 1 \end{pmatrix}.$$

5. *The Pseudo-orthogonal Groups.* The definition of $O(n)$ as the group of isometries of the dot product of $\mathbb{R}^n$

$$O(n) = \{A \in GL(n, \mathbb{R}) \mid \langle A\cdot, A\cdot \rangle = \langle \cdot, \cdot \rangle\}$$

can be generalized by considering isometries of other bilinear forms.

Here are a couple of examples. Consider the (not positive definite) inner product of $\mathbb{R}^{p+q}$ given by

$$\langle u, v \rangle_{(p,q)} := -u_1 v_1 - \cdots - u_p v_p + u_{p+1} v_{p+1} + \cdots + u_{p+q} v_{p+q}.$$

The *pseudo-orthogonal group* of signature $p - q$ is defined as

$$O(p, q) = \{A \in GL(p+q, \mathbb{R}) \mid \langle A\cdot, A\cdot \rangle_{(p,q)} = \langle \cdot, \cdot \rangle_{(p,q)}\}.$$

When $p = 1$ and $q = 3$ we obtain the *Lorentz group* of relativity theory.

6. *The Symplectic Group.* If instead of a symmetric bilinear form we take the alternating bilinear form in $\mathbb{R}^{2n}$ given by

$$\omega(u, v) := (u_1 v_{n+1} - u_{n+1} v_1) + (u_2 v_{n+2} - u_{n+2} v_2) + \cdots + (u_n v_{2n} - u_{2n} v_n)$$

then we obtain the group

$$Sp(2n, \mathbb{R}) = \{A \in GL(2n, \mathbb{R}) \mid \omega(A\cdot, A\cdot) = \omega(\cdot, \cdot)\}$$

called the *symplectic group*. Both the pseudo-orthogonal and the symplectic group can be shown to be Lie groups by the method used above for $O(n)$. This is also a consequence of the following fundamental fact, which will not be proved here.

THEOREM 4.0.29. *Any closed subgroup of a Lie group is also a Lie group.*

Many real Lie groups are more naturally defined as subgroups of the space of complex matrices: $M(n, \mathbb{C})$. For example, $GL(n, \mathbb{C})$ can be regarded as a real group by means of the homomorphism

$$A = A_1 + iA_2 \in GL(n, \mathbb{C}) \mapsto \begin{pmatrix} A_1 & -A_2 \\ A_2 & A_1 \end{pmatrix} \in GL(2n, \mathbb{R})$$

where A_1 and A_2 have real entries. This allows us to characterize groups such as the *unitary group*

$$U(n) = \{A \in GL(n, \mathbb{C}) \mid A^* A = I\}$$

also as real Lie groups. (Here A^* indicates transpose followed by complex conjugation of all entries of the matrix.) Note that $U(n) \cap GL(n, \mathbb{R}) = O(n)$.

g. Rigid motions in $\mathbb{R}^3$. Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a diffeomorphism of $\mathbb{R}^n$, that is, an invertible everywhere differentiable map whose inverse is also everywhere differentiable. We say that f is an *isometry* of $\mathbb{R}^n$ if it preserves distances. More precisely, for any two points $p, q \in U$,

$$\operatorname{dist}(f(p), f(q)) = \operatorname{dist}(p, q),$$

where $\operatorname{dist}(p, q) := |p - q|$ denotes the Euclidean distance between p and q.

We show next that the isometries of $\mathbb{R}^n$ are all affine maps whose linear part is in $O(n)$.

LEMMA 4.0.30. For each $p \in U$, the (Jacobian matrix of the) differential df_p of an isometry $f : U \to \mathbb{R}^n$ is an element of $O(n)$.

PROOF. Note the following:

$$\begin{aligned} |df_p u| &= \left| \lim_{h\to 0} \frac{f(p+hu) - f(p)}{h} \right| \\ &= \lim_{h\to 0} \left| \frac{f(p+hu) - f(p)}{h} \right| \\ &= \lim_{h\to 0} \frac{|f(p+hu) - f(p)|}{h} \\ &= \lim_{h\to 0} \frac{|p+hu-p|}{|h|} \\ &= |u|. \end{aligned}$$

Therefore, df_p is an orthogonal transformation. □

PROPOSITION 4.0.31. Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be an isometry of $\mathbb{R}^n$. Fix any $p_0 \in \mathbb{R}^n$ and define a map $T : \mathbb{R}^n \to \mathbb{R}^n$ by $T(p) := f(p+p_0) - f(p_0)$. Then, T is an element of $O(n)$, and it does not depend on the choice of p_0.

PROOF. For each $u \in \mathbb{R}^n$, $|u| = 1$, define the curve $\gamma(t) = T(tu)$. Since T is a diffeomorphism sending 0 to 0, we are allowed to write for any $t \neq 0$

$$v(t) := T(tu)/|T(tu)|$$

so that $\gamma(t) = |t|v(t)$. For $t > 0$ we have $\gamma'(t) = v(t) + tv'(t)$. Moreover, since $v(t)$ is a unit vector we have $\langle v', v\rangle = 0$. Also, for $t > 0$, $v'(t) = -T(tu)/t^2 + dT_{tu}u/t$, so that

$$|\gamma'(t)|^2 = |v(t)|^2 + t^2|v'(t)|^2 = 1 + t^2 \left| -\frac{T(tu)}{t^2} + \frac{dT_{tu}u}{t} \right|^2.$$

Using that $|\gamma'(t)| = |dT_{tu}u| = |u| = 1$, it follows that $T(tu)/t = dT_{tu}u$. Therefore, for all $v \in \mathbb{R}^n$,

$$\langle u, v\rangle = \langle \frac{T(tu)}{t}, T(v)\rangle = \langle dT_{tu}u, T(v)\rangle = \langle \gamma'(t), T(v)\rangle.$$

Differentiating in t, we obtain $\langle \gamma''(t), T(v)\rangle = 0$ for $t > 0$. Since T is a diffeomorphism, it follows that $\gamma''(t) = 0$, so $\gamma'(t)$ is a constant vector. Therefore, $T(tu) = wt$, for some vector w. Differentiating in t at $t = 0$ we get $w = dT_0 u$. Consequently, $T(u) = Au$ for all u, where $A = dT_0 \in O(n)$. Therefore, there is $q_0 \in \mathbb{R}^n$ such that $f(p) = q_0 + Ap$ for all p. Written in this form, it is clear that df_p does not depend on p. □

It follows from the proposition that any isometry of $\mathbb{R}^n$ is of the form

$$f(p) = f(0) + Ap$$

where A is an element of $O(n)$.

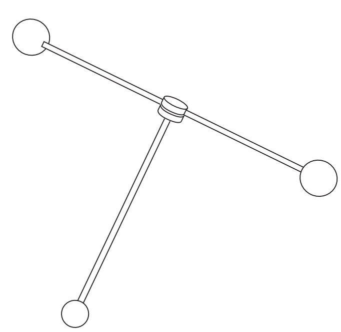
h. Motion of Rigid Bodies. Consider a solid body, free to move in $\mathbb{R}^3$. A *reference configuration* $\mathcal{B}$ of the body is the closure of an open set in $\mathbb{R}^3$ with a piecewise smooth boundary. A *configuration* of $\mathcal{B}$ is a mapping $\sigma : \mathcal{B} \to \mathbb{R}^3$ (say, C^1, invertible on its image, orientation preserving). Given $x \in \mathcal{B}$, the image $\sigma(x)$ is the location in space occupied by x.

We say that $\mathcal{B}$ is a *rigid body* if its configurations are isometries of $\mathbb{R}^3$.

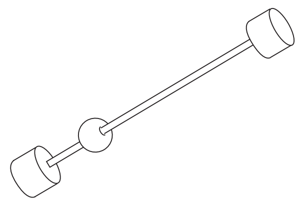
The group $SE(3)$ acts transitively on the set of all configurations of a rigid body, so that this set can be regarded as a homogeneous space of the form $SE(3)/H$, where H is the group of symmetries of the body.

A *motion* of the rigid body $\mathcal{B}$ (during the time interval $[t_0, t_1]$) is a continuous path $t \in [t_0, t_1] \mapsto \sigma(t) \in SE(3)$.

The following picture represents a system of masses connected by rigid rods freely moving in space. It's space of configurations is the quotient of $SE(3)$ by a finite group of symmetries. (Strictly speaking, for it to be consider a rigid body the connection between the two perpendicular rods should not be allowed to rotate. If the mass represented by the smaller sphere is regarded as a point mass and the rod to which it is attached is regarded as being 1-dimensional, then the definition of rigid body applies.)



The mechanical system described in the next picture also affords the same number of degrees of freedom as the previous one. It consists of a rigid rod moving freely in space, along which can slide a rigid spherical bead. After a doubling procedure similar to what we used for the billiard table, it is possible to describe its configuration manifold as $\mathbb{R}^3 \times S^2 \times \mathbb{T}^1$. Notice that $SO(3)$ and $S^2 \times \mathbb{T}^1$ are not diffeomorphic.



EXERCISE 4.0.32. Show that $SO(3)$ and $S^2 \times \mathbb{T}^1$ are not diffeomorphic. (If you know something about fundamental groups this is easy: $SO(3)$ is diffeomorphic to $\mathbb{R}P^3$, which is S^3 modulo the equivalence relation that identifies antipode points. From this it can be shown that the fundamental group of $SO(3)$ is the finite group on two elements. On the other hand, the fundamental group of $S^2 \times \mathbb{T}^1$ is the product of the fundamental group of S^2, which is trivial, and of $\mathbb{T}^1$, which is $\mathbb{Z}$. Since the two manifolds have non-isomorphic fundamental groups, they cannot be diffeomorphic.

i. Group Actions and Homogeneous Spaces. An *action* of a group G on a set M is defined by a map

$$\Phi : G \times M \to M$$

having the following two properties: for each $p \in M$, $\Phi(e, p) = p$, where e is the identity element in G; and for all $g - 1, g_2 \in G$ and all $p \in M$,

$$\Phi(g_1 g_2, p) = \Phi(g_1, \Phi(g_2, p)).$$

This can be rephrased by saying that the map $g \mapsto \Phi_g := \Phi(g, \cdot)$ is a homomorphism from G into the group of bijection from M to itself. We can also say that the action represents G as a group of transformations of M.

If G is a Lie group and M is a differentiable manifold, we can demand that an action be differentiable, that is, that Φ be a differentiable map.

The *orbit* of $p \in M$ under the action is a subsect of M that consists of all $\Phi(g, p)$, $g \in G$. We can define an equivalence relation on M be declaring that two points are equivalent if their orbits are the same. The set of equivalence classes will be denoted by M/G. Often the quotient topology and quotient measurable structure of such spaces have bad saparation properties, e.g., they may fail to be Hausdorff. However, the following result shows that if M is itself a Lie group and G is a closed subgroup acting on M by left (or right) translations, then M/G has a natural differentiable manifold structure. The theorem will be stated without proof. (We change notations in the theorem to reflect how the symbols are likely to appear later in the text.)

Theorem 4.0.33. *Let H be a closed subgroup of a Lie group G, and Φ the action of H on G given by $\Phi(h,g) = hg$. Then G/H carries a unique differentiable structure such that:*

(1) *The projection $\pi : G \to G/H$ that associates to each g its orbit is differentiable.*
(2) *For every $p \in G/H$ there exists an open neighborhood W of p and a differentiable map $\phi : W \to G$ such that $\pi \circ \phi$ is the identity map on G/H.*

Also, left-translations $L_{g_0} : gH \mapsto g_0 gH$, are diffeomorphisms from G/H to itself.

We say that a group action (of G on M) is *transitive* if M is a single orbit. In other words, for each $p \in M$, the map $f : g \in G \mapsto gp \in M$ is surjective. If this is the case, let H denote the *isotropy subgroup* of p, which is the subgroup $\{g \in G : gp = p\}$. This is a closed subgroup of G (if the action is at least continuous). It is now possible to define a bijection between G/H and M that associates an equivalence class gH to each $gH \in G/H$. (Verify that this correspondence is well-defined, and that it is indeed a bijection.) By identifying M with G/H, M acquires a structure of differentable manifold.

We give now a few examples of homogeneous spaces.

1. *Sphere.* The group $SO(n+1)$ acts transitively on the unit sphere S^n in $\mathbb{R}^{n+1}$ via the action $(A, x) \mapsto Ax$, where x is viewed as a column vector. The isotropy group of $(0, \ldots, 0, 1)$ is isomorphic to $SO(n)$. Therefore we can view S^n as a homogeneous space $SO(n+1)/SO(n)$. In fact, it is not difficult to show that $SO(n+1)/SO(n)$ (with the quotient manifold structure) and S^n (regarded as a submanifold of $\mathbb{R}^{n+1}$) are diffeomorphic.

2. *Projective Space.* The general linear group $GL(n, \mathbb{R})$ acts transitively on the set of 1-dimensional subspaces of $\mathbb{R}^n$ under the action that maps $(A, [u]) \mapsto [Au]$, where u is a nonzero vector in $\mathbb{R}^n$ and $[u]$ represents the line spanned by u. (Of course $SO(n)$ already acts transitively there.) The isotropy subgroup of the line $\{(0, \ldots, t) : t \in \mathbb{R}\}$ is easily seen to be the subgroup H of $GL(n, \mathbb{R})$ consisting of matrices of the form

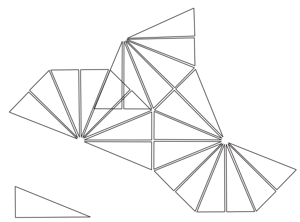
$$\begin{pmatrix} * & \ldots & * & 0 \\ \vdots & & \vdots & \vdots \\ * & \ldots & * & 0 \\ * & \ldots & * & * \end{pmatrix}.$$

Therefore, we can regard $\mathbb{R}P^{n-1}$ as a homogeneous space $GL(n, \mathbb{R})/H$.

3. *The n-torus.* A discrete subgroup of a Lie group is a 0-dimensional Lie group. We have already seen that the n-torus is a homogeneous space by writing it as $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$.

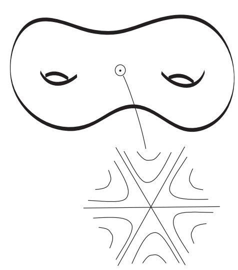
Exercise 4.0.34 (Positive Quadratic Forms). Show that the space of all positive definite (symmetric) quadratic forms on $\mathbb{R}^n$ can be identified with the quotient $SL(n, \mathbb{R})/SO(n)$.

j. Surfaces and Billiards. The idea of describing the motion of a billiard ball on a rectangular table by straight lines on a torus can be extended to more general shapes of tables. For example, consider a table in the shape of a right triangle with one of the angles equal to $\pi/8$. Try to convince yourself that the motion of the billiard ball on the triangle can also be described by considering straight line motions on a compact flat surface obtained by glueing together 16 copies of the triangle (or its mirror image).



Topologically, what one gets is a closed (compact, without boundary) surface of genus 2. It is also a smooth manifold on the complement of one singular point. The singular point is the center of a disc of total angle 6π. A neighborhood of the singular point can be parametrized by a neighborhood of the origin of the Riemann surface of the cube root of z.

The singularity is unavoidable if one wants a genus 2 (or greater) surface built out of flat pieces. A very similar construction, but now yielding a perfectly regular surface, is possible by considering billiard tables on the Poincaré disc, which is a model of a surface of constant negative curvature. We will return to such things later in our discussion of Riemannian geometry.



1. Tangent Vectors

Since differentiable manifolds are not in general vector spaces, the notion of a tangent vector at a given point of a manifold M has to be defined with some care.

If we think of a tangent vector as something that describes the velocity of a differentiable curve, and that the velocity of the curve determines the rate of change of any (differentiable) function on the curve, then it seems natural to make the following definition: Two differentiable curves, $\alpha_i : (-a, a) \to M$, $i = 1, 2$, passing through a point $p = \alpha_i(0)$, will be said to have the same velocity at that point if for every differentiable function $f : U \to \mathbb{R}$, where $U \subset M$ is a neighborhood of p, the real valued functions $f \circ \alpha_1$ and $f \circ \alpha_2$ have the same first derivative at $t = 0$. Having the same velocity at p is an equivalence relation on the set of differentiable curves (defined on some interval of 0) that pass through p at time 0. The velocity vector can be regarded as the equivalence class itself. We say that v is a *tangent vector to M at p* if it is the velocity vector of some differentiable curve passing through p.

Of course, for such a definition to be valid, we need to make sure that, at least on submanifolds of $\mathbb{R}^n$, this is nothing but the ordinary notion of vector. If M is a submanifold of $\mathbb{R}^n$, then by a simple application of the chain rule two curves α_1, α_2 such that $\alpha_1(0) = \alpha_2(0)$ have the same derivative at $t = 0$ if and only if $(f \circ \alpha_1)'(0) = (f \circ \alpha_2)'(0)$ for all real valued differentiable function f. So in this case nothing new is gainned and nothing old is lost.

The set of tangent vectors at p will be denoted T_pM. Given a tangent vector v at p, and a differentiable function defined on a neighborhood of p, the *directional derivative* of f on the direction of v will be written

$$D_v f = vf := \frac{d(f \circ \alpha)}{dt}(0)$$

for every differentiable curve with velocity v at p. As the notation $vf = D_v f$ suggests, we will often identify the vector itself with its operation of directional derivative. In that spirit, we will often write expressions such as $a_1 \frac{\partial}{\partial x_1} + \cdots + a_n \frac{\partial}{\partial x_n}$ to represented a vector.

Exercise 4.1.1. Show that If M is an n-dimensional manifold, then T_pM is an n-dimensional vecor space for all $p \in M$.

Let $f : M \to N$ be a differentiable map between two differentiable manifolds. The differential of f at a point p, already discussed for submanifolds of $\mathbb{R}^n$, easily extends to general manifolds. It is a linear map

$$df_p : T_pM \to T_{f(p)}N$$

defined as follows: If $u \in T_pM$ and u is the velocity vector at p of a differentiable curve α, then $df_p u$ is the vector in $T_{f(p)}N$ corresponding to the curve $f \circ \alpha$.

Exercise 4.1.2. Let G be a closed subgroup of $GL(n, \mathbb{R})$ and denote by $F : G \to G$ the operation $F(g) = g^{-1}$. Show that for all $g \in G$ and $v \in T_gG$ we have $dF gv = -g^{-1}vg^{-1}$. (Compare this with the derivative $\frac{d}{dt}(x^{-1}) = -x'/x^2$.)

Exercise 4.1.3. If $M = S^n = \{x \in \mathbb{R}^{n+1} : |x| = 1\}$, show that

$$TS^n = \{(x, v) \in \mathbb{R}^{n+1} \times \mathbb{R}^{n+1} : |x| = 1, x \cdot v = 0\}.$$

Use the implicit function theorem to show that TS^n is a differentiable submanifold of $\mathbb{R}^{2n+2}$.

Exercise 4.1.4. Let G be a closed subgroup of $GL(n, \mathbb{R})$ of the form

$$G = \{g \in GL(n, \mathbb{R}) : gJg^t = J\},$$

where J is a fixed element of $M(n, \mathbb{R})$. Show that the tangent space T_eG at the identity element of G is the subspace $\mathfrak{g} = \{V \in M(n, \mathbb{R}) : VJ + JV^t = 0\}$.

2. The Tangent Bundle

Let M be a differentiable manifold and let

$$TM = \{(p, v) : p \in M, v \in T_pM\}.$$

Let $\pi : TM \to M$ the natural projection $(p, v) \mapsto p$. A vector field V on M is a function from M to TM such that $\pi \circ V$ is the identity map of M. We can give TM a manifold structure such that π and is a differentiable map and a vector field V, regarded as a map in the way just described, is differentiable if and only if for every differentiable function $f : M \to \mathbb{R}$, Vf is also a differentiable function (of one degree less than f). TM has dimension $2n$ if M has dimension n.

The differentiable structure of TM can be described using the following system of parametrizations. Let $\phi : U \to M$ be a differentiable parametrization of an open set $U \subset \mathbb{R}^n$. Define $\Phi : U \times \mathbb{R}^n \to TM$ by setting $\Phi(p, v) := (\phi(p), d\phi_p v)$. If ψ is another parametrization of U and $\Psi : U \times \mathbb{R}^n \to TM$ is the corresponding local parametrization of TM then the change of parametrizations is $(p, v) \mapsto (\psi^{-1}(\phi(p)), d(\psi^{-1} \circ \phi)_p v)$, which is differentiable (of degree one less than ϕ and ψ.) Notice, in fact, that we obtain this way more than a manifold structure on TM. The change of parametrization maps have the special property that for each p, their restriction to $\{p\} \times \mathbb{R}^n$ maps linearly onto $\{q\} \times \mathbb{R}^n$, $q = \psi^{-1}(\phi(p))$. Therefore the linear space structure of $\mathbb{R}^n$ can be coherently transferred to each T_pM. This makes TM a *vector bundle*, called the *tangent bundle of* M.

A differentiable *vector field* on M is a differentiable map $V : M \to TM$ such that $V(p) \in T_pM$ for each p. Since we will be regarding tangent vectors as derivative operations on functions, it will often be convenient to write v_p rather than $v(p)$, so that the directional derivative of some f along v is denoted $v_p f$.

1. *Example.* If $f : U \subset \mathbb{R}^n \to M$ is a differentiable parametrization of M and V is a vector field on M, then $f_*^{-1}V = df^{-1}V \circ f$ is the local expression of V in the coordinates defined by f. As an example, let V be the vector field on $S^3 \subset \mathbb{R}^4$ such that

$$V(x) = \begin{pmatrix} x_2 \\ -x_1 \\ x_4 \\ -x_3 \end{pmatrix}.$$

We would like to find the coordinate expression of V for the parametrization

$$\phi : y \mapsto (y, \sqrt{1 - |y|^2})$$

where $y \in B_1 = \{y \in \mathbb{R}^3 : |y| < 1\}$. In other words, we would like to find functions f_1, f_2, f_3 on B_1 such that

$$\phi_*^{-1}V = f_1 \frac{\partial}{\partial y_1} + f_2 \frac{\partial}{\partial y_2} + f_3 \frac{\partial}{\partial y_3}.$$

Exercise 4.2.1. Show that

$$\phi_* \frac{\partial}{\partial y_i} = \frac{\partial}{\partial x_i} - \frac{x_i}{x_4} \frac{\partial}{\partial x_4}$$

for $i = 1, 2, 3$.

It follows from the previous exercise that

$$\phi_*^{-1}V = y_2 \frac{\partial}{\partial y_1} - y_1 \frac{\partial}{\partial y_2} + \sqrt{1 - |y|^2} \frac{\partial}{\partial y_3}.$$

Exercise 4.2.2. If X is a vector field on M, $F : M \to N$ is a diffeomorphism and $f : N \to \mathbb{R}$ is a differentiable function, then

$$(F_* X)f = X(f \circ F).$$

Exercise 4.2.3. Let $X_1, \ldots, X_k$ be C^∞ vector fields on M. We can regard $X_1 \ldots X_k$ as a differential operator of order k on functions. If $F : M \to N$ is a C^∞ diffeomorphism and $f : N \to \mathbb{R}$ is a C^∞ function, then

$$(X_1 \ldots X_k)(f \circ F) = (F_* X_1) \ldots (F_* X_k) f.$$

Exercise 4.2.4. If X and Y are differentiable vector fields on $\mathbb{R}^n$, show that

$$[X, Y] := XY - YX$$

is a derivative operator of first order. In fact $[X, Y]$ is itself a vector field.

Exercise 4.2.5. Use the previous exercises to show that the commutator of two vector fields X, Y on a manifold M is also a vector field. (In particular, it is a derivative operator of first order.) Do this by showing that

$$[X, Y] = \phi_* [\phi_*^{-1} X, \phi_*^{-1} Y],$$

for any local parametrization ϕ of M.

3. Vector fields and flows

A differentiable vector field on M is a differentiable map $V : M \to TM$ such that $V(p) \in T_p M$ for all $p \in M$. If V is a differentiable vector field and f a differentiable function on M, then Vf is a differentiable function on M (of degree one less than f).

a. Vector Fields and Differential Equations. To a vector field V on M one associates a differential equation

$$\dot{p} = V(p).$$

A solution of such equation is a path γ in M such that $\gamma'(t) = V(\gamma(t))$ for each t. Conversely, systems of ordinary differential equations correspond to vector fields.

For example, consider the initial value problem

$$\ddot{x} = f(t, x, \dot{x})$$

with initial conditions $x(0) = x_0$, $x'(0) = \dot{x}_0$. By introducing coordinates $x_1 = t, x_2 = x, x_3 = \dot{x}$, the given equation is equivalent to the system

$$\begin{aligned} \dot{x}_1 &= 1 \\ \dot{x}_2 &= x_3 \\ \dot{x}_3 &= f(x_1, x_2, x_3). \end{aligned}$$

These may be regarded as the components of a vector field

$$V(x) = \frac{\partial}{\partial x_1} + x_3 \frac{\partial}{\partial x_2} + f(x) \frac{\partial}{\partial x_3}.$$

A solution $x(t)$ of the initial value problem is a solution of the differential equation such that $x(0) = (0, x_0, \dot{x}_0)$.

Exercise 4.3.1 (Bernoulli Equation). The Bernoulli differential equation is the first order non-linear equation $y' + P(x)y = f(x)y^n$.

(1) Show that the Bernoulli equation is equivalent to the system

$$\dot{y}_1 = 1$$
$$\dot{y}_2 = -P(y_1)y_2 + f(y_1)y_2^n,$$

which defines the vector field

$$V(y_1, y_2) = \frac{\partial}{\partial y_1} + (-P(y_1)y_2 + f(y_1)y_2^n)\frac{\partial}{\partial y_2}.$$

(2) Suppose that $n \neq 1$ and define $(w_1, w_2) = \phi(y_1, y_2)$, where $w_1 = y_1, w_2 = y_2^{1-n}$. Show that

$$W := \phi_* V = \frac{\partial}{\partial w_1} + (1-n)(-P(w_1)w_2 + f(w_1))\frac{\partial}{\partial w_2}.$$

(3) Show that the vector field W comes from a first order, inhomogeneous differential equation

$$\frac{1}{1-n}\frac{dw}{dx} + P(x)w = f(x).$$

Exercise 4.3.2 (Riccati Equation). Consider the equation

$$\frac{dy}{dx} = \frac{y^2}{x^3} + \frac{y+1}{x}.$$

The graphs of solutions to this equation are flow lines of the vector field

$$V(x, y) = \frac{\partial}{\partial x} + (\frac{y^2}{x^3} + \frac{y+1}{x})\frac{\partial}{\partial y}.$$

(1) Show that if Φ is a local diffeomorphism of the plane such that $\Phi_* V = hV$ for any real valued function h, then Φ sends the graph of a solution of the initial equation to the graph of another solution. In other words, Φ is a *symmetry* of the differential equation.
(2) Let Φ_t be a one-parameter family of symmetries with generator X, a vector field on the plane. Explain that $[X, V]$ must be of the form lV, where l is a real valued function. A vector field having the latter property will be called an *infinitesimal symmetry* of the differential equation.
(3) Let $X = \alpha\frac{\partial}{\partial x} + \beta\frac{\partial}{\partial y}$ be an infinitesimal symmetry of the differential equation and suppose that α and β are polynomial functions of x and y of degree less than or equal to 2. Show that

$$\alpha = cx^2,\ \beta = cxy.$$

(4) Find a one-parameter group of symmetries of the given differential equation. Do this by finding the flow of the X obtained previously. This amounts to solving a simple differential equation. The result should be

$$\Phi_t(x, y) = (\frac{x}{1-tx}, \frac{y}{1-tx}).$$

Check by direct computation that Φ_t is indeed a one-parameter group.
(5) Let Ψ_s be an arbitrary flow on the plane having infinitesimal generator Y such that X and Y are not colinear. For example, pick $\Psi_s(x, y) = (x, y + s)$. Fix a point $p_0 = (x_0, y_0)$ and define

$$\varphi : (s, t) \mapsto \Phi_t(\Psi_s(p_0)).$$

Show that φ defines the coordinate change (local diffeomorphism) on the plane: writing $p_0 = (x_0, y_0)$,

$$t = \frac{1}{x_0} - \frac{1}{x}$$
$$s = x_0 \frac{y}{x} - y_0.$$

(6) Show that $\varphi_* \frac{\partial}{\partial t} = X$ and that $\frac{\partial}{\partial t}$ is an infinitesimal symmetry of the differential equation in the new coordinate system. (The effect of using the flow of X to define new coordinates is that, the field of slopes associated to the vector field $\varphi_*^{-1} V$ is invariant under the flow of $\frac{\partial}{\partial t}$, which consists of translations on the t-direction. Therefore, this field of slopes does not depend on t.)

(7) Show that in the new coordinate system the original equation becomes (for simplicity we take here $x_0 = 1, y_0 = 0$):

$$\frac{ds}{dt} = s^2 + 1.$$

This equation is easily solved:

$$s = \tan(t - C).$$

Going back to x, y-coordinates, we obtain the solution

$$y = x \tan(C - \frac{1}{x})$$

for the original equation.

b. Lie Brackets. Let U be an open subset of $\mathbb{R}^n$. A vector field on $\mathbb{R}^n$ can also be expressed as

$$v = \sum_{i=1}^{n} f_i \frac{\partial}{\partial x_i}.$$

This means that for each p, v_p is the derivation at p given by

$$v_p g = \sum_{i=1}^{n} f_i(p) \frac{\partial g}{\partial x_i}(p).$$

The derivative notation makes it easier to introduce one operation on vector fields that will turn out to be very useful, called the Lie bracket. It is defined as follows. Given two C^{k+1} vector fields $v = \sum_{i=1}^n f_i \frac{\partial}{\partial x_i}$ and $w = \sum_{i=1}^n g_i \frac{\partial}{\partial x_i}$ with domain U, the bracket $[v, w]$ is the C^k vector field on U defined as follows. Let $p \in U$ and let h be any C^2 function on some neighborhood of p. Then

$$[v, w]_p h = v_p(wh) - w_p(vh).$$

In other words, the bracket corresponds to the commutator of the two derivations. In principle, an expression like the one defining $[v, w]$ could be a second order derivation. That $[v, w]_p$ is really a vector at p (i.e., a first order derivation) follows from the calculation

$$v_p(wh) - w_p(vh) = \sum_{i=1}^{m} f_i(p) \left(\frac{\partial}{\partial x_i} (\sum_{j=1}^{m} g_j \frac{\partial h}{\partial x_i}) \right)_p - \sum_{i=1}^{m} g_i(p) \left(\frac{\partial}{\partial x_i} (\sum_{j=1}^{m} f_j \frac{\partial h}{\partial x_i}) \right)_p$$
$$= \sum_{i,j=1}^{m} \left(f_i(p) \frac{\partial g_j}{\partial x_i}(p) - g_i(p) \frac{\partial f_j}{\partial x_i}(p) \right) \frac{\partial h}{\partial x_j}(p)$$

It was used above that $\frac{\partial^2 h}{\partial x_i \partial x_j}(p) = \frac{\partial^2 h}{\partial x_j \partial x_i}(p)$.

EXERCISE 4.3.3. Show that the bracket of vector fields satisfies the following properties. If u, v, w are smooth vector fields, a, b are real constants, and f is a smooth function, then

(1) $[v, w] = -[w, v]$
(2) $[au + bv, w] = a[u, w] + b[v, w]$
(3) $[u, fv] = (uf)v + f[u, v]$
(4) $[u, [v, w]] = [[u, v], w] + [v, [u, w]]$

The last of the properties enumerated in the exercise is called the *Jacobi identity.* A vector space with a bilinear operation that satisfies properties 1,2, and 4 is called a *Lie algebra*. The above exercise says that the space of smooth vector fields on $\mathbb{R}^m$ is a Lie algebra with respect to the bracket operation.

By viewing vector fields X, Y on a manifold M as derivative operators (of order 1) on functions, it makes sense to consider the (Lie bracket) operator (a priori of order 2) defined by

$$[X, Y]f = X(Yf) - Y(Xf)$$

where f is any C^2 function. This is, in fact, a vector field, which can be checked in local coordinates using the previous discussion. The linear space of C^∞ vecor fields on a manifold M forms a Lie algebra under the Lie bracket operation.

1. *Linear Vector Fields.* As an example we consider vector fields on $\mathbb{R}^n$ of the form $u(p) = -Ap$ where $A = (a_{ij})$ is an $n \times n$ real matrix. In terms of the basis $\{\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}\}$, u can be written as follows:

$$u(p) = \sum_{ij} a_{ij} x_j(p) \frac{\partial}{\partial x_i}.$$

(We write $x_i : \mathbb{R}^n \to \mathbb{R}$ for the i-th coordinate function.)

EXERCISE 4.3.4. Given A, B two $n \times n$ matrices, show that

$$[u_A, u_B] = u_{[A,B]}$$

where $[A, B] = AB - BA$. (The commutator of two matrices.)

DEFINITION 4.3.5. Let v be a vector field. An *integral curve* of v with initial condition p is a differentiable path $c : J \to U$, defined on an open interval J containing 0, such that $c(0) = p$ and $c'(t) = v(c(t))$ for all $t \in J$.

We visualize the field v by associating the vector $v(x) \in \mathbb{R}^n$ to x in such a way that the initial point of the vector lies at x rather than at the origin of $\mathbb{R}^n$. The velocity vector of an integral curve of v at x is precisely $v(x)$.

Note that finding an integral curve through a point p for a vector field v exactly corresponds to the problem of finding a solution to a system of ordinary differential equations, often written in the form $\dot{x_i} = v_i(x)$, $i = 1, \dots, n$, with initial condition

$x(0) = p$. We recall now the main facts about existence and uniqueness of solutions of systems of ODEs, and about the dependence of solutions on initial conditions.

THEOREM 4.3.6 (Existence-uniqueness theorem). Let U be an open subset of $\mathbb{R}^n$ and $v : U \to \mathbb{R}^n$ a C^1 vector field. For any given $p \in U$, there exists an integral curve $\lambda : (-a, a) \to U$ of v with the initial condition $c(0) = p$. If $\mu : (-b, b) \to U$ is another integral curve of v with $\mu(0) = p$, then $\lambda = \mu$ on some interval $(-c, c)$ contained in the intersection of $(-a, a)$ and $(-b, b)$.

PROOF. Let B be a closed ball with center p and radius r, over which the norms $|v(x)|$ and $|dv_x|$ are bounded by a number k for all $x \in B$. In particular, an application of the mean value theorem yields

$$|v(x) - v(y)| \leq k|x - y|$$

for all $x, y \in B$. Let a be a positive real number such that $ak < \min\{1, r\}$. We define a metric space E consisting of the continuous paths $\lambda : [-a, a] \to B$ with the metric of uniform convergence. Recall that E is complete. We define a map $\mathcal{F} : E \to E$ by

$$\mathcal{F}(\lambda)(t) := p + \int_0^t v(\lambda(s))\ ds.$$

It is immediate that

$$|\mathcal{F}(\lambda(t)) - p| \leq ak < r$$

so that $\mathcal{F}(\lambda)(t) \in B$, hence $\mathcal{F}(\lambda) \in E$. Also note that

$$\begin{aligned} |\mathcal{F}(\lambda_1)(t) - \mathcal{F}(\lambda_2)(t)| &\leq \int_0^t |v(\lambda_1(s)) - v(\lambda_2(s))|\ ds \\ &\leq ak\sup\{|\lambda_1(s) - \lambda_2(s)| : -a \leq s \leq a\}. \end{aligned}$$

Therefore

$$\mathrm{dist}(\mathcal{F}(\lambda_1), \mathcal{F}(\lambda)_2) \leq ak\mathrm{dist}(\lambda_1, \lambda_2).$$

Since $ak < 1$, we conclude that $\mathcal{F}$ is as contraction. By the fixed point theorem for contractions there exists a unique path $\lambda : [-a, a] \to B$ such that $\mathcal{F}(\lambda) = \lambda$. But this means that

$$\lambda(t) = p + \int_0^t v(\lambda(s))\ ds.$$

By the fundamental theorem of calculus, this is equivalent to

$$\lambda'(t) = v(\lambda(t)), \quad \lambda(0) = p.$$

Uniqueness of the integral curve is a consequence of uniqueness of the fixed point for contractions. □

It is worth recording the following consequence of the above proof.

PROPOSITION 4.3.7. Suppose that v is a C^1 vector field defined on the open subset $U \subset \mathbb{R}^n$. Let $p \in U$ and suppose that for all q in a closed ball $B \subset U$ of radius r and center p the numbers $|v(q)|$ and $|dv_q|$ are bounded above by a number k. Then there is an integral curve $t \mapsto \lambda(t)$ of v defined over an interval $[-a, a]$ so long as $a < \min\{1/k, r/k\}$. In particular, if v is defined on all of $\mathbb{R}^n$ such that $|v(q)|$ and $|dv_q|$ are bounded by a constant independent of q, then there is $a > 0$ such that for every point p we can find an integral curve of v with initial point q that is defined over $[-a, a]$.

EXERCISE 4.3.8. Given any $A \in M(n, \mathbb{R})$ we can define a vector field v on all of $\mathbb{R}^n$ by $v(x) = Ax$. Show that the unique integral curve to v passing through a point $x_0 \in \mathbb{R}^n$ is given by $\gamma(t) = e^{tA}x_0$. (The exponential is defined by the power series $\sum_{i=0}^{\infty} 1/n!(tA)^n$.) Find the matrix e^{tA} explicitly for the following cases:

(1) $A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$

(2) $A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 0 & -1 \end{pmatrix}$.

We describe now some refinements of the existence and uniqueness theorem, mostly stated without proof. The main conclusions will be summarized in the next theorem. Let U and v be as in the theorem. It can be shown that for each $p \in U$, there exists an interval $J \subset \mathbb{R}$ containing 0 and an integral curve $\lambda : J \to U$, $\lambda(0) = p$, such that J is maximal in the following sense: If $\mu : I \to U$ is another integral curve defined on an interval I containing 0 satisfying $\mu(0) = p$, then $I \subset J$ and $\mu(t) = \lambda(t)$ for all $t \in I$. We call J the *maximal interval of existence*. It turns out that J is an open interval and has the following continuity property: If $[a, b] \subset J$ is a closed interval containing 0 in its interior, then for every q in a sufficiently small open neighborhood of p there is an integral curve with initial point q that is defined on $[a, b]$.

For each $p \in U$, let $I(p)$ denote the maximal interval of existence for the initial point p. The unique integral curve defined on $I(p)$ will be denoted $\Phi_t(p)$. Suppose that v is C^r. The function

$$\Phi : \Omega := \{(t, p) \in \mathbb{R} \times U \mid t \in I(p)\} \to U$$

is also C^r and, for each p, the function of t given by $t \mapsto \Phi_t(p)$ C^{r+1}.

For each $p \in U$, $t \in I(p)$, and $s \in I(\Phi_t(p))$, we have $t + s \in I(p)$ and

$$\Phi_{t+s}(p) = \Phi_s(\Phi_t(p)).$$

We can see this for $s > 0$ as follows. Write $I(p) = (a, b)$ and define the function

$$\lambda(r) = \begin{cases} \Phi_r(p) & \text{if } a < r \leq t \\ \Phi_{r-t}(\Phi_t(p)) & \text{if } t \leq r \leq s + t \end{cases}.$$

Then λ is an integral curve for v on $(a, s + t]$ with initial point p, so $s + t \in I(p)$. Moreover, by uniqueness of solutions $\Phi_{t+s}(p) = \lambda(s + t) = \Phi_s(\Phi_t(p))$ as claimed. The case $s = 0$ is trivial and the case $s < 0$ is shown in a similar way by defining

$$\lambda(r) = \begin{cases} \Phi_r(p) & \text{if } t \leq r < b \\ \Phi_{r-t}(\Phi_t(p)) & \text{if } s + t \leq r \leq t \end{cases}.$$

It can also be shown that If $(t, p) \in \Omega$, there is an open neighborhood V of p such that $\Phi_t(V)$ is open and $\Phi_t : V \to \Phi_t(V)$ is a C^r diffeomorphism. Using the comments of the previous paragraph we obtain that $\Phi_t^{-1} = \Phi_{-t}$. More precisely, for all $p \in V$, and all $q \in \Phi_t(V)$ we have

$$\Phi_{-t}(\Phi_t(p)) = p, \quad \Phi_t(\Phi_{-t}(q)) = q.$$

The next theorem summarizes the comments just made.

THEOREM 4.3.9. Let v be a C^r vector field on an open set $U \subset \mathbb{R}^n$. For each $p \in U$ let $I(p)$ denote the maximal interval of existence and $t \mapsto \Phi_t(p)$ the unique integral curve defined on $I(p)$. Then there exists an open subset $\Omega \subset \mathbb{R} \times U$ such that for each $p \in U$

$$I(p) = \{t \in \mathbb{R} \mid (t, p) \in \Omega\},$$

for all $t \in I(p)$ and $s \in I(\Phi_t(p))$ we have that $s + t \in I(p)$, and

$$\Phi_{s+t}(p) = \Phi_s(\Phi_t(p)).$$

Each Φ_t, $t \in I(p)$, is a diffeomorphism from an open neighborhood of p onto an open neighborhood of $\Phi_t(p)$, whose inverse is given by Φ_{-t}. Finally, for each $p \in U$ and for each interval $[a, b] \subset I(p)$ having 0 in its interior, there exists an open neighborhood N of p such that $[a, b] \subset I(q)$ for all $q \in N$.

A map Φ satisfying the properties given in the theorem is called a *local flow*. When $I(p) = \mathbb{R}$ for each p and Φ_t is a diffeomorphism from U onto itself, we say that Φ is a *global flow* or simply a *flow*.

For example, the linear vector field $p \mapsto Ap$, for $p \in \mathbb{R}^n$ and $A \in M(n, \mathbb{R})$, is a vector field whose (global) flow is

$$\Phi_t(p) = e^{tA}p.$$

The homomorphism property defining a flow corresponds in this case to:

$$e^{(t+s)A} = e^{tA}e^{sA}, \quad e^{-tA} = (e^{tA})^{-1}.$$

EXERCISE 4.3.10. Consider the vector field on $U = (0, \infty) \subset \mathbb{R}$ given by $v(x) = (1/x)e$. ($e = \frac{\partial}{\partial x}$ is the unit basis vector of $\mathbb{R}$.) Find the local flow associated to v and the maximal intervals of existence. Check that the defining properties of a local flow are satisfied for this example.

EXERCISE 4.3.11. Repeat the previous problem for the vector field on $\mathbb{R}^2$ defined by $v(x, y) = (-x, y + x^2)$. Draw a sketch of the flow lines.

PROPOSITION 4.3.12. If v is a C^1 vector field on $\mathbb{R}^n$ such that $|v(q)|$ and $|dv_q|$ are bounded above by a constant k independent of q, then v admits a global flow.

PROOF. We saw above that if $t \in I(p)$ and $s \in I(\Phi_t(p))$, then $t+s \in I(p)$. But proposition 1.3.4 implies that $I(q)$ contains an interval $[-a, a]$ for a positive a that is independent of $q \in \mathbb{R}^n$. Therefore $I(p)$ must actually contain $[-na, na]$ for an arbitrary n, so that $I(p) = \mathbb{R}$ for all p. It is now easy to show, using the uniqueness of solutions, that the Φ_t are global diffeomorphisms and we have a global flow. $\square$

We would like now to relate the operation of taking the bracket with a vector field v on the one hand and the flow of X on the other. Towards this goal, we first need to understand how vector fields transform under a diffeomorphism. So let U and V be open subsets of $\mathbb{R}^n$ and $f : U \to V$ a C^∞ diffeomorphism. Let v be a C^k vector field on U.

DEFINITION 4.3.13. The *push-forward* of v under f is the C^k vector field f_*v on V defined by

$$(f_*v)_p := df_{f^{-1}(p)}v_{f^{-1}(p)}$$

for every $p \in V$.

EXERCISE 4.3.14. Let $f = (f_1, \ldots, f_m)$ be a diffeomorphism from U to V as above. Show that for every smooth function $h : V \to \mathbb{R}$, we have

$$(f_*v)_p h = v_{f^{-1}(p)}(h \circ f).$$

When h is the coordinate function x_j and v is the standard basis element $\frac{\partial}{\partial x_i}$, conclude that

$$(f_* \frac{\partial}{\partial x_i})_p x_j = \frac{\partial f_j}{\partial x_i}(f^{-1}(p))$$

so that

$$f_* \frac{\partial}{\partial x_i} = \sum_{j=1}^{n} \left(\frac{\partial f_j}{\partial x_i} \circ f^{-1} \right) \frac{\partial}{\partial x_j}.$$

The diffeomorphism f may be regarded in two ways. Either as a transformation of $\mathbb{R}^m$ that moves points around, or as a change of coordinates. We can make the latter viewpoint more clearly if we give a different name to the coordinates on V, say y_i, and continue to call x_i the coordinates on U. In this form, $y_i = f_i(x_1, \ldots, x_n)$ for $i = 1, \ldots, n$. For example, let V be the complement in $\mathbb{R}^3$ to the half-plane $\{(x, y, z) \in \mathbb{R}^3 \mid y = 0, x \geq 0\}$ and let U be the open box $(0, \infty) \times (0, 2\pi) \times (0, \pi)$, with coordinates (ρ, θ, φ). The spherical coordinates on $\mathbb{R}^3$ are given by the following f:

$$\begin{aligned} x &= f_1(\rho, \theta, \varphi) = \rho \sin\varphi \cos\theta \\ y &= f_2(\rho, \theta, \varphi) = \rho \sin\varphi \sin\theta \\ z &= f_3(\rho, \theta, \varphi) = \rho \cos\varphi. \end{aligned}$$

EXERCISE 4.3.15. Find an expression for $f_* \frac{\partial}{\partial \rho}$ in the coordinates (x, y, z).

DEFINITION 4.3.16. We denote by $\mathcal{L}_v$ the operation on vector fields defined by

$$\mathcal{L}_v w := [v, w].$$

$\mathcal{L}_v$ is called the *Lie derivative* of vector fields with respect to v.

Note that the Lie derivative indeed behaves like a derivative in the following two ways: Let u, v, w be smooth vector fields on $\mathbb{R}^n$ and h a smooth function from $\mathbb{R}^n$ into $\mathbb{R}$. Then

$$\mathcal{L}_u(hv) = (uh)v + h\mathcal{L}_u v$$

where hv is the vector field such that $(hv)_p := h(p)v_p$, and

$$\mathcal{L}_u[v, w] = [\mathcal{L}_u v, w] + [v, \mathcal{L}_u w].$$

The second identity is simply the Jacobi identity, and both are part of an earlier exercise. Although we have defined the Lie derivative for vector fields over $\mathbb{R}^n$ it is clear that it also makes sense for vector fields over an open set $U \subset \mathbb{R}^n$.

The next proposition says, in poetic language, that $\mathcal{L}_v$ is the "infinitesimal generator" for the action of the flow of v on vector fields. It makes clear in what way the Lie derivative is a derivation.

PROPOSITION 4.3.17. Let v and w be a smooth vector fields on $\mathbb{R}^n$ and let the local flow of v be $\Phi_t : U \to \mathbb{R}^n$, where $U \subset \mathbb{R}^n$ is open. Then

$$(\mathcal{L}_v w)_p = \left(\frac{d}{dt} (\Phi_{-t*} w)_p \right)_{t=0}$$

PROOF. We first note that if h is any smooth function, and since the flow of v depends smoothly on t, there is $r(p,t)$ such that $r(p,t)/t$ goes to 0 with t, such that

$$h(\Phi_t(p)) = h(p) + tv_p h + r(p,t)t.$$

We now have:

$$\begin{aligned}\left(\frac{d}{dt}\left(\Phi_{-t*}w\right)_p h\right)_{t=0} &= \left(\frac{d}{dt}w_{\Phi_t(p)}(h\circ\Phi_{-t})\right)_{t=0}\\ &= \left(\frac{d}{dt}w_{\Phi_t(p)}(h - tvh + r(\cdot,-t)t)\right)_{t=0}\\ &= \left(\frac{d}{dt}(wh)(\Phi_t(p))\right)_{t=0} - \left(\frac{d}{dt}tw_{\Phi_t(p)}(vh)\right)_{t=0}\\ &= v_p(wh) - w_p(vh)\\ &= [v,w]_p h.\end{aligned}$$

□

EXERCISE 4.3.18. Consider the map $\Phi_t : \mathbb{R}^n \to \mathbb{R}^n$ defined by $\Phi_t(p) = e^t p$. Show that Φ is the (global) flow associated to the vector field $v(p) = p$. Using the derivative notation for vector fields, show that $v = \sum_{i=1}^n x_i \partial/\partial x_i$. Also show that

$$\left(\Phi_{t*}\frac{\partial}{\partial x_i}\right)_p = e^t\left(\frac{\partial}{\partial x_i}\right)_p.$$

Conclude that if w is a constant vector field (i.e., $w = \sum_{i=1}^n a_1 \frac{\partial}{\partial x_i}$, where the a_i are constants) then

$$\left(\frac{d}{dt}(\Phi_{-t*}w)_p\right)_{t=0} = -w.$$

Verify by direct computation that $[v,w] = -w$.

EXERCISE 4.3.19. Show that if f is a diffeomorphism of $\mathbb{R}^n$ and v and w are smooth vector fields, then

$$f_*[v,w] = [f_*v, f_*w].$$

4. More on vector fields

Let now M be a C^k submanifold of $\mathbb{R}^n$ of dimension m. A vector field on M is a map $v : M \to \mathbb{R}^n$. It is a C^r field if v is a C^r function. It is a *tangent vector field* on M if $v(p) \in TM_p$ for each $p \in M$. For the most part, vector fields on M will be tangent fields, so we normally omit the adjective. Sometimes we may consider vector field defined only on some open set $U \subset M$. We then refer to it as a vector field on U.

Given a parametrization φ of $U \subset M$, the vectors $\frac{\partial\varphi}{\partial x_i}(x) = d\varphi_x e_i$ form for each $x \in U_0$ a basis of TM_p, $p = \varphi(x)$. The functions $\frac{\partial\varphi}{\partial x_i} : U_0 \to \mathbb{R}^n$, $i = 1,\dots,m$, are C^{k-1} and so $v_i(\varphi(x)) = \frac{\partial\varphi}{\partial x_i}(x)$ are vector fields of class C^{k-1} on U. These vector fields constitute the *moving frame* associated to the parametrization.

EXERCISE 4.4.1. Fix $r \le k-1$. A (tangent) vector field $v : M \to \mathbb{R}^n$ is of class C^r if and only if for each parametrization $\varphi : U_0 \to U$ of class C^k and each

$p = \varphi(x) \in U$ we have

$$v(p) = \sum_{i=1}^{m} \alpha_i(x) \frac{\partial \varphi}{\partial x_i}(x)$$

where the functions $\alpha_i : U_0 \to \mathbb{R}$ are of class C^r.

We now extend to submanifolds of $\mathbb{R}^n$ the basic facts about integrating vector fields. Let $v : M \to \mathbb{R}^n$ a tangent vector field (from now on, simply a vector field) on M. An *integral curve* of v is a differentiable path $\gamma : (-\epsilon, \epsilon) \to M$ with $\gamma(0) = p$ and $\gamma'(t) = v(\gamma(t))$, for all t.

PROPOSITION 4.4.2. Let v be a C^{k-1}, $k \geq 2$, vector field on a C^k submanifold $M \subset \mathbb{R}^n$ of dimension m. Then for each point $p \in M$ there exists an integral curve of v in M with initial point p. Two integral curves of v with origin p must coincide on a neighborhood of 0.

PROOF. Given $p \in M$, let $\varphi : U_0 \to U \subset M$be a parametrization of class C^k of a neighborhood U of p in M. We define a vector field $u : U_0 \to \mathbb{R}^n$ of class C^k by the condition:

$$d\varphi_x u(x) = v(\varphi(x))$$

for all $x \in U_0$. The chain rule shows that $\mu : (-\epsilon, \epsilon) \to U_0$ is an integral curve of u with origin $p_0 = \varphi^{-1}(p)$ if and only if $\varphi \circ \mu : (-\epsilon, \epsilon) \to U$ is an integral curve of v with origin $p = \varphi(p_0)$. The proposition then follows by the corresponding theorem for vector fields on $\mathbb{R}^m$. □

COROLLARY 4.4.3. Let $W \subset \mathbb{R}^r$ be an open set, $M \subset W$ a C^k, $k \geq 2$, manifold of dimension m and $v : W \to \mathbb{R}^r$ a vector field of class C^{k-1} on W such that for all $p \in M$, $v(p) \in TM_p$. If $\lambda : (-\epsilon, \epsilon) \to \mathbb{R}^r$ is an integral curve of v with initial point $p \in M$, then there exists $\delta > 0$ such that $\lambda(t) \in M$ for all t for which $|t| < \delta$.

PROOF. In fact, the restriction of v to M is a vector field on M of class C^k. By the proposition, for every $p \in M$ there exists an integral curve of v, with initial point p, contained in M. By uniqueness, this curve is the restriction of λ to a neighborhood of 0. □

COROLLARY 4.4.4. If M is a compact submanifold of $\mathbb{R}^n$ and v is a C^1 vector field on M, then v admits a global flow.

EXERCISE 4.4.5. Suppose that $M \subset \mathbb{R}^n$ and $N \subset \mathbb{R}^k$ are smooth submanifolds with same dimension m and let $f : M \to N$ be a smooth diffeomorphism. If v is a C^∞ vector field on M, it is possible to define a vector field w on N by $w(p) = df_{f^{-1}(p)} v(f^{-1}(p))$, called the *push-forward* of v to N. Show that if Φ_t denotes the local flow of v, then $\Psi_t := f \circ \Phi_t \circ f^{-1}$ is the local flow associated to w.

EXERCISE 4.4.6. Show that the vector field $v(x_1, x_2, x_3, x_4) = (-x_3, x_4, x_1, -x_2)$ is tangent to the sphere S^3 at every point $x \in S^3$. Obtain an explicit expression for the global flow on S^3 associated to v. Show that the flow on S^3 is the restriction to the sphere of the linear flow $\Phi_t(p) = e^{At} p$ on $\mathbb{R}^4$, where

$$A = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0. \end{pmatrix}$$

Since the matrix is skew-symmetric, $\Phi_t \in SO(4)$ for each t. In particular, Φ_t indeed maps the sphere to itself.

EXERCISE 4.4.7 (Differential equations on $\mathbb{R}P^n$). Let $\pi : \mathbb{R}^{n+1} - \{0\} \to \mathbb{R}P^n$ be the natural projection, so that $\pi(p)$ is the line spanned by p. For each $s \neq 0$ and $p \in \mathbb{R}^{n+1}$, let $F_s(p) = sp$. We will say that a vector field V on $\mathbb{R}^{n+1} - \{0\}$ is a radial vector field if

$$V(p) = l(p) \sum_{i=1}^{n+1} x_i(p) \frac{\partial}{\partial x_i},$$

for some function l and all $p \in \mathbb{R}^{n+1} - \{0\}$.

(1) Show that if Z is a vector field on $\mathbb{R}P^n$, then there exists a vector field X on $\pi : \mathbb{R}^{n+1} - \{0\}$ such that $d\pi_p X = Z_{\pi(p)}$ for all $p \in \mathbb{R}^{n+1} - \{0\}$. Show that X must satisfy the property that $(F_s)_* X - X$ is radial. Conversely, if a vector field X on $\mathbb{R}^{n+1} - \{0\}$ is such that $(F_s)_* X - X$ is radial, then it projects under π to a vector field on $\mathbb{R}P^n$. A vector field X of this type will be called *projectable*.

(2) Show that all linear vector fields are projectable. Recall that a linear vector field has the form $X(p) = \sum_{i,j=1}^{n+1} a_{ij} x_j(p) \frac{\partial}{\partial x_i}$.

(3) Let U denote the unit radial vector field pointing away from the origin. Given a vector field X, set $\bar{X} = X - \langle X, U \rangle U$. Here $\langle X, U \rangle$ denotes the ordinary dot product, hence $\bar{X}$ is the component of X that is tangent to the sphere of radius $|p|$. Show that X is projectable if and only if $(F_s)_* \bar{X} = \bar{X}$ for all $s \neq 0$.

(4) Show that vector field X on S^n is projectable to $\mathbb{R}P^n$ if and only if $(F_{-1})_* X = X$. If $\Phi_t(p)$ is the flow of X, then show that the flow of $\pi_* X$ is given by Ψ_t such that $\Psi_t(\pi(p)) = \pi(\Phi_t(p))$ for every $p \in S^n$.

Redo everything here without assuming that the manifolds are embedded.

5. Lie algebras of matrix groups

A subgroup $G \subset GL(n, \mathbb{R})$ is called a (matrix) Lie group if it is a C^∞ submanifold of $M(n, \mathbb{R})$. We have already seen a few examples: $GL(n, \mathbb{R})$ itself, $SL(n, \mathbb{R})$, $O(n)$ and $SO(n)$. Lie groups of matrices are also called *linear* Lie groups.

Given a linear Lie group $G \subset GL(n, \mathbb{R})$, the tangent space TG_I to G at the identity I is called the *Lie algebra* of G. It has indeed a (nonassociative) algebra naturally attached to it, which may be viewed as describing the group structure of G "infinitesimally." This algebra is described next.

Given two n by n matrices A and B define their *commutator* (or *Lie bracket*) by

$$[A, B] = AB - BA.$$

The operation $(A, B) \mapsto [A, B]$ is a bilinear map from $M(n, \mathbb{R}) \times M(n, \mathbb{R})$ into $M(n, \mathbb{R})$. It is *alternating* (or *anti-symmetric*): $[A, B] = -[B, A]$, and satisfied the *Jacobi identity* :

$$[A, [B, C]] + [C, [A, B]] + [B, [C, A]] = 0.$$

These properties are all immediate from the definition of $[\cdot, \cdot]$ and their verification is left to the reader.

(It may be easier to remember the Jacobi identity by writing it in the form

$$[A, [B, C]] = [[A, B], C] + [B, [A, C]].$$

In this form, it is apparent that taking the bracket with A is an algebraic derivation with respect to the Lie bracket operation, i.e., the Jacobi identity is nothing but Leibnitz's product rule for differentiation.)

Let $\mathfrak{g} \subset M(n, \mathbb{R})$ be a vector subspace such that for all $A, B \in \mathfrak{g}$ the bracket $[A, B]$ is also in $\mathfrak{g}$. We say in this case that $\mathfrak{g}$ is a *Lie algebra of matrices.*

Of course, $M(n, \mathbb{R})$ is itself a Lie algebra of matrices. The same is true for the trivial subspace consisting of 0. A less trivial example is

$$\mathfrak{o}(n) = \{A \in M(n, \mathbb{R}) : A^t = -A\}.$$

EXERCISE 4.5.1. Show that $\mathfrak{o}(n)$, just defined, is a Lie algebra.

We now show that the tangent space at the identity of a linear Lie group G is a Lie algebra. Towards this goal we define the *exponential* of a matrix $A \in M(n, \mathbb{R})$ as

$$e^A = I + A + \frac{1}{2!}A^2 + \cdots + \frac{1}{n!}A^n + \dots$$

It is shown in basic linear algebra courses that the series always converges and that for all commuting A, B(i.e., such that $[A, B] = 0$) we have

$$e^A e^B = e^{A+B}.$$

In particular, $e^{(s+t)A} = e^{sA}e^{tA}$ and $e^A e^{-A} = e^0 = I$, so that e^A is invertible for all A and $(e^A)^{-1} = e^{-A}$.

Differentiating term by term, we find

$$\frac{d}{dt}(e^{tA}) = Ae^{tA}.$$

In particular, $f : \mathbb{R} \to GL(n, \mathbb{R})$ defined by $f(t) = e^{tA}$ is a C^∞ path whose velocity vector at $t = 0$ is A. A more precise result is the following.

PROPOSITION 4.5.2. Let $G \subset M(n, \mathbb{R})$ be a Lie group of matrices. For any $A \in T_I G$, we have $e^{tA} \in G$ for all $t \in \mathbb{R}$.

PROOF. Consider the vector field $v : GL(n, \mathbb{R}) \to M(n, \mathbb{R})$ defined by $v(X) = AX$. The path $\lambda : \mathbb{R} \to M(n, \mathbb{R})$ defined by $\lambda(t) = e^{tA}X$ is an integral curve of v with initial point X. Notice that $v(X) = (dR_X)_I A$, the image of A under the differential of the right-multiplication by X. Since right-multiplication by an element of G restricts to a diffeomorphism of G onto itself, then for $X \in G$ we have $V(X) \in TG_X$. It follows that for each $X \in G$ there is $\epsilon > 0$ such that $e^{tA}X \in G$ whenever $|t| < \epsilon$. In particular, taking $X = I$, we have $e^{tA} \in G$ for $|t| < \epsilon$. For an arbitrary t, write $t = t_1 + \cdots + t_k$ so that $|t_i| < \epsilon$ and notice that each $e^{t_i A} \in G$. Since G is a group, the product $e^{t_1 A} \dots e^{t_k A} = e^{tA}$ is also in G. □

PROPOSITION 4.5.3. Let G be a Lie group of matrices. Given $A, B \in TG_I$, we have $[A, B] \in TG_I$. In other words, the tangent space to G at I is a Lie algebra.

PROOF. For every $t \in \mathbb{R}$ write $\alpha(t) = e^{tA}$ and $\beta(t) = e^{tB}$. By the previous proposition we have $\alpha(t) \in G$ and $\beta(t) \in G$. Therefore, we may define a path $\lambda : [0, \infty) \to G$, given by

$$\lambda(t) = \alpha(\sqrt{t})\beta(\sqrt{t})\alpha(\sqrt{-t})\beta(\sqrt{-t}).$$

Write

$$\alpha(t) = I + tA + \frac{t^2A^2}{2} + \rho(t), \quad \beta(t) = I + tB + \frac{t^2B^2}{2} + \sigma(t),$$

where the limit of $\rho(t)/t^2$ and $\sigma(t)/t^2$ as $t \to 0$ are 0. Then by a simple calculation, we obtain

$$\lambda(t) = I + t[A, B] + \theta(t),$$

where $\theta(t)/t$ goes to 0 as $t > 0$ goes to 0. Therefore, $\lambda'(0) = [A, B]$. Since $\lambda(t) \in G$ for all $t \geq 0$, we conclude that $[A, B] \in TG_I$. □

Note that for any $X \in \mathfrak{g} := T_I G$, the set of matrices of the form e^{tX} form a subgroup of G. The map $\sigma(t) = e^{tX}$ is referred to as a 1-*parameter subgroup* of G.

a. A Little More on Isometries of Euclidian Space. We now study in greater detail the group of isometries of $\mathbb{R}^n$. For each $a = (a_1, a_2, a_3) \in \mathbb{R}^3$, define the matrix

$$\omega(a) = \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix}$$

Denote by $\mathfrak{so}(3)$ the Lie algebra of $SO(3)$ (which is the same as the Lie algebra of $O(3)$). Recall that the *cross-product* on $\mathbb{R}^3$ is the (nonassociatice) product defined by

$$(a_1, a_2, a_3) \times (b_1, b_2, b_3) = (a_2 b_3 - b_2 a_3, b_1 a_3 - a_1 b_3, a_1 b_2 - b_1 a_2).$$

EXERCISE 4.5.4. Show that $(\mathbb{R}^3, [\cdot, \cdot])$, with bracket $[a, b] := a \times b$, is a Lie algebra isomorphic to $\mathfrak{so}(3)$. Show this by proving that

$$\omega(a \times b) = [\omega(a), \omega(b)]$$

where the bracket on the right-hand side is the commutator of matrices. In particular, this shows that the cross-product satisfies the Jacobi identity

$$a \times (b \times c) + b \times (c \times a) + c \times (a \times b) = 0.$$

Also show that $\omega(a)b^t = (a \times b)^t$, where $a, b \in \mathbb{R}^3$ and c^t indicates the transpose of a row vector c.

EXERCISE 4.5.5. For all $a, b \in \mathbb{R}^3$, show that

$$\omega(a)^2 b = \langle a, b \rangle a - |a|^2 b$$
$$\omega(a)^3 b = -|a|^2 \omega(a) b.$$

By using $e^{\theta\omega(a)} = \sum_{n=0}^{\infty} \frac{\theta^n}{n!} \omega(a)^n$, conclude that

$$e^{\omega(a)} b = b + (a \times b) \sin\theta + (\langle a, b \rangle a - |a|^2 b)(1 - \cos\theta).$$

EXERCISE 4.5.6. Let $a \in \mathbb{R}^3$ be a unit vector and let b, c be unit vectors such that $\langle a, b, c \rangle$ form a positive orthonormal basis of $\mathbb{R}^n$. Denote $R(a, \theta) := e^{\theta\omega(a)} \in SO(3)$. Show that

$$R(a, \theta) a = a.$$

Using the previous exercise, show that

$$R(\theta, a) b = b \cos\theta + c \sin\theta.$$

Conclude that $R(\theta, a)$ corresponds to a rotation about the axis determined by a, so that the plane perpendicular to a is rotated by an angle θ.

EXERCISE 4.5.7. Show that for any $S \in SO(3)$, $a \in \mathbb{R}^3$, and $\theta \in \mathbb{R}$, we have

$$R(\theta, Sa) = SR(\theta, a)S^{-1}.$$

Interpret the result geometrically. Use this to show that

$$S(a \times b) = Sa \times Sb$$

for all $a, b \in \mathbb{R}^3$.

By what was seen before, every element of $\mathfrak{so}(3)$ is of the form $\theta\omega(a)$, $|a| = 1$. Therefore, the next proposition says that every element of $SO(3)$ is of the form $R(\theta, a)$ for some θ and a. In particular, every element of $SO(3)$ is a rotation about some axis (one-dimensional subspace) in $\mathbb{R}^3$.

PROPOSITION 4.5.8. The exponential map $\exp : \mathfrak{so}(3) \to SO(3)$, $\exp(A) := e^A$, is surjective.

PROOF. Let $S \in SO(3)$. We claim that 1 is an eigenvalue of S. Note that S has a nonzero real eigenvalue λ, since its characteristic polynomial has degree 3. Let a be a unit eigenvector associated to λ. Since $\lambda^2|a|^2 = \langle Sa, Sa\rangle = |a|^2$, λ must be either 1 or -1. If the other eigenvalues are nonreal, then we have two complex conjugate eigenvalues, μ and $\bar{\mu}$, so that $1 = \det(A) = \lambda\mu\bar{\mu} = \lambda|\mu|^2$, from which we deduce that $\lambda = 1$. Notice that S restricts to an orthogonal transformation on the plane $a^\perp$ perpendicular to a. If the transformation $S|_{a^\perp}$ has one real eigenvalue with unit eigenvector b, then a unit vector perpendicular to both a and b will also be an eigenvector. Therefore, if $\lambda = -1$, S is diagonalizable with real eigenvalues, which are all either 1 or -1. But the product of the eigenvalues is 1, so at least one of them must be 1, showing the claim.

Let a be a unit eigenvector associated to eigenvalue 1. Let b and c be unit vectors such that $\langle a, b, c\rangle$ forms a positive orthogonal basis of $\mathbb{R}^3$. Then $Sb = \alpha b + \beta c$, where α and β are real numbers such that $\alpha^2 + \beta^2 = 1$. Therefore, there is $\theta \in \mathbb{R}$ such that $\alpha = \cos\theta$ and $\beta = \sin\theta$. Since $\langle Sa, Sb, Sc\rangle$ is also a positive basis (since the determinant of S is 1), we must have that $Sc = -\beta b + \alpha c$. Therefore, $S = R(\theta, a)$. □

We now study in a similar way the full group of orientation preserving isometries, $SE(3)$. We regard $SE(3)$ as the subgroup of $GL(4, \mathbb{R})$ consisting of matrices of the form

$$\begin{pmatrix} A & u \\ 0 & 1 \end{pmatrix}$$

where $A \in SO(3)$ and $u \in \mathbb{R}^3$.

The 1-parameter subgroups of $SE(3)$ are discussed next. Recall that these are the maps (group homomorphisms from $\mathbb{R}$ into G) of the form $\sigma(t) = \exp(t\omega)$, for some $\omega \in \mathfrak{se}(3)$ (the Lie algebra of $SE(3)$).

EXERCISE 4.5.9. Show that the Lie algebra of $SE(n)$ is the set of matrices in $M_{n+1}(\mathbb{R})$ of the form

$$\begin{pmatrix} X & w \\ 0 & 0 \end{pmatrix}$$

where $X \in \mathfrak{so}(n)$ and w is any vector in $\mathbb{R}^n$. Check that

$$\begin{pmatrix} X & w \\ 0 & 0 \end{pmatrix}^n = \begin{pmatrix} X^n & X^{n-1}w \\ 0 & 0 \end{pmatrix}$$

and use this to show that

$$\sigma(t) := \exp\left(t\begin{pmatrix} X & w \\ 0 & 0 \end{pmatrix}\right) = \begin{pmatrix} e^{tX} & \left(\int_0^t e^{sX} ds\right) w \\ 0 & 1 \end{pmatrix}.$$

Show by direct computation that $\sigma(t)$ satisfies the homomorphism property

$$\sigma(t_1 + t_2) = \sigma(t_1)\sigma(t_2)$$

that characterizes an exponential.

Conclude that if $X = \omega(a)$, for $a \in \mathbb{R}^3$ such that $|a| = 1$,

$$\sigma(\theta) = \begin{pmatrix} R(\theta, a) & (I - R(\theta, a))(a \times w) + \langle a, w\rangle a\theta \\ 0 & 1 \end{pmatrix}.$$

(Note: You will need the identity $a \times (a \times w) = \langle a, w\rangle a - |a|^2 w$.)

PROPOSITION 4.5.10. The exponential map $\exp : \mathfrak{se}(3) \to SE(3)$ is surjective.

PROOF. First note that

$$\exp\left(\begin{pmatrix} 0 & u \\ 0 & 0 \end{pmatrix}\right) = \begin{pmatrix} I & u \\ 0 & 1 \end{pmatrix}.$$

Therefore, we may assume that $A \in SE(3)$ has the form

$$A = \begin{pmatrix} S & u \\ 0 & 1 \end{pmatrix}$$

where $S \in SO(3)$ is not equal to the identity I. We already know that $S = R(\theta, a)$ for some unit vector $a \in \mathbb{R}^3$ and some $\theta \in (0, 2\pi)$. Given the previous exercise, it suffices to show that the transformation $T_a : \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$T_a : w \mapsto (I - R(\theta, a))(a \times w) + \langle a, w\rangle a\theta$$

is nonsingular. Let b and c be unit vectors such that $\langle a, b, c\rangle$ form a positive orthonormal basis of $\mathbb{R}^3$. Then

$$\begin{aligned} R(\theta, a)a &= a \\ R(\theta, a)b &= b\cos\theta + c\sin\theta \\ R(\theta, a)c &= -b\sin\theta + c\cos\theta. \end{aligned}$$

Moreover, for any $S \in SO(3)$, an easy calculation gives $ST_aS^{-1} = T_{Sa}$. Therefore, we may assume without loss of generality that $a = e_1$, $b = e_2$, $c = e_3$. Then

$$\begin{aligned} T_{e_1}e_1 &= \theta e_1 \\ T_{e_1}e_2 &= \sin\theta e_2 + (1 - \cos\theta)e_3 \\ T_{e_1}e_3 &= (-1 + \cos\theta)e_2 + \sin\theta e_3. \end{aligned}$$

The matrix associated to T_{e_1} in the standard basis is

$$\begin{pmatrix} \theta & 0 & 0 \\ 0 & \sin\theta & 1 - \cos\theta \\ 0 & -1 + \cos\theta & \sin\theta \end{pmatrix}.$$

Since the determinant of the above matrix is $2\theta(1 - \cos\theta) \neq 0$, T_{e_1} is nonsingular, concluding the proof. □

EXERCISE 4.5.11. Find the matrix in $SE(3) \subset GL(4,\mathbb{R})$ that corresponds to rotation by $\pi/4$ in the positive direction about the axis that passes through the point $(1,2,3)$, parallel to the vector from the origin to the point $(1,1,1)$.

We show now that any 1-parameter subgroup of $SE(3)$ may be described as a *screw motion.* A screw S consists of an *axis* l, a *pitch* h, and a *magnitude* M. A screw motion represents rotation by an angle $M = \omega t$ about the axis l followed by translation by an amount $(h\omega/2\pi)t$ parallel to the axis l. A pure rotation corresponds to $h = 0$ and a pure translation corresponds to $\omega = 0$.

Note that, if l is the line through a point p_0 parallel to a vector a, then a screw motion is given by the following rigid transformation:

$$f(p) = p_0 + R(\theta, a)(p - p_0) + \frac{h\theta}{2\pi} a$$

where $\theta = \omega t$. As an element of $GL(4,\mathbb{R})$, f corresponds to the matrix

$$\begin{pmatrix} R(\theta, a) & (I - R(\theta, a))p_0 + \frac{h\theta}{2\pi} a \\ 0 & 1 \end{pmatrix}.$$

But we know that any element $g \in SE(3)$ can be written as

$$\begin{pmatrix} R(\theta, a) & (I - R(\theta, a))(a \times v) + \langle a, v\rangle a\theta \\ 0 & 1 \end{pmatrix}.$$

Therefore, we see that g corresponds to a screw if we choose $p_0 = a \times v$ and $h = \langle a, v\rangle 2\pi$.

b. Rolling. Consider a sphere that rolls over a flat surface without slipping or twisting. In this section we would like to describe the *phase space* of such a system.

Let S^2 denote the unit sphere in $\mathbb{R}^3$ with center at the origin and let $\mathbb{R}^2$ be the plane embedded in $\mathbb{R}^3$ as the subset for which the third coordinate of each of its points is 0. We would like to represent the motion of a sphere rolling over the plane by a curve in $SE(3)$ such that, for each time t, the "rolling" sphere is the image $g(t)S^2$ of the "model" sphere (which remains fixed at the origin) under an isometry $g(t)$. The conditions that the sphere touches the plane tangentially at one point and that it does not slip or twist will have to be imposed by appropriate equations that restrict the values of $g(t)$ and its derivative at each moment of time t.

Our physical intuition suggests that given a differentiable path $\eta(t)$ on the plane $\mathbb{R}^2$, there should exist a unique path $t \mapsto g(t) \in SE(3)$ such that the moving sphere $t \mapsto g(t)S^2$ rolls over the plane "without slipping and twisting" in such a way that $\eta(t)$ is the point of contact of the moving sphere with the plane, for each time t. We will show below that this intuition indeed holds for the geometric model of the rolling sphere that we now describe.

We first introduce the *rolling condition*. The rolling of S^2 over $\mathbb{R}^2$ along a path $t \mapsto \eta(t) \in \mathbb{R}^2$ is defined by a path $t \mapsto g(t) \in SE(3)$ such that

$$\gamma(t) := g(t)^{-1}\eta(t)$$

is a path in S^2 and $dg(t)_{\gamma(t)}$ (which is just the element in $SO(3)$ corresponding to the "rotation part" of $g(t)$) sends $T_{\gamma(t)}S^2$ isomorphically onto $T_{\eta(t)}\mathbb{R}^2 = \mathbb{R}^2$.

It will be useful to define a manifold P whose points describe the possible configurations of the moving sphere that are allowed by the rolling condition. First note that a configuration of the sphere can be specified by a triple (p, g, q), where $p \in S^2$, $q \in \mathbb{R}^2$ and $g = (A, u) \in SE(3)$, satisfying the properties: $gp = q$ and dg_p

maps T_pS^2 isomorphically onto $T_q\mathbb{R}^2 = \mathbb{R}^2$. The second condition is equivalent to that $A = dg_p$ should map the unit normal vector to S^2 at p, which is simply p itself, to the unit normal vector $-e_3$ to $\mathbb{R}^2$, i.e., $Ap = -e_3$. (The negative sign is due to our viewing the sphere as rolling on the upper part of the plane rather than on its underside.) The first condition can be expressed by $Ap + u = q$ or, as $Ap = -e_3$, by $u = q + e_3$. Therefore, an allowed configuration of the moving sphere becomes completely determined by specifying an element (A, q) in $P := SO(3) \times \mathbb{R}^2$ and each such element describes an allowed configuration. In fact, from each (A, q) we can recover $p = -Ae_3$ and $g = (A, q + e_3)$. We denote the natural projection onto the second factor by $\pi : P \to \mathbb{R}^2$. The rolling of S^2 over the curve $\eta(t)$ in $\mathbb{R}^2$ is then completely specified by a path $(A(t), \eta(t))$ in P.

We now define the no-slip and no-twist conditions. Suppose that a curve $g(t) \in SE(3)$ describes a rolling of the unit sphere over the curve $\eta(t)$ on the plane, and write $\gamma(t) = g(t)^{-1}\eta(t)$. For a fixed t, the curve

$$s \mapsto c(s) := g(t+s)\gamma(t), \quad s \in (-\epsilon, \epsilon),$$

describes the motion of $\eta(t) \in S^2$ immediately before and after it touches the plane. At the precise moment $s = 0$ when it touches the plane the point should be stationary: $c'(0) = 0$. In other words, we define the *no-slip condition* by $g'(t)\gamma(t) = 0$ for each t.

The following example will make the condition clear. Consider a disc rolling on a line so that for each t the length traveled along the line equals the arclength spanned by the moving point. The trajectory of a point at the border of the disc can then be parametrized by the curve $c(t) = (t - \sin t, 1 - \cos t)$. The curve c is differentiable and $c'(0) = 0$, which is the no-slip condition in this case.

In order to define the no-twist condition, consider for each t the curve

$$s \in (-\epsilon, \epsilon) \mapsto g(t+s)g(t)^{-1} \in SE(3).$$

If v is a tangent vector to the plane at the point $\eta(t)$, then

$$v(s) := d(g(t+s)g(t)^{-1})_{\eta(t)}v$$

gives the motion of v during a small time before and after the moment of contact at $\eta(t)$. The *no-twist condition* is that $v'(0)$ should not have nonzero component tangent to the plane:

$$d(g'(t)g(t)^{-1})_{\eta(t)}v \in (\mathbb{R}^2)^\perp$$

for all $v \in \mathbb{R}^2 = T_{\eta(t)}\mathbb{R}^2$.

We now show how the no-slip and no-twist conditions can be expressed in terms of the manifold P. First, write a tangent vector to P at (A, q) as $(\dot{A}, \dot{q})$. Here $\dot{A} = A'(0)$ and $\dot{q} = q'(0)$, for curves $A(t)$ and $q(t)$ in $SO(3)$ and $\mathbb{R}^2$, respectively. Keeping in mind that $(A(t), q(t)$ determines $(p(t), g(t), q(t)) \in S^2 \times SE(3) \times \mathbb{R}^2$ by $p(t) = -A(t)^{-1}e_3$ and $g(t) = (A(t), q(t) + e_3)$, the no-slip condition $g'(0)p(0)$ translates to $\dot{q} = \dot{A}A^{-1}e_3$. For the no-twist condition, note that $d(g'(0)g(0)^{-1})_q = \dot{A}A^{-1}$, so $\dot{A}A^{-1}e_1$ and $\dot{A}A^{-1}e_2$ do not have non-zero tangential component along $\mathbb{R}^2$.

Writing $\dot{q} = (\dot{x}, \dot{y})$ and

$$\dot{A}A = \begin{pmatrix} 0 & -\alpha_3 & \alpha_2 \\ \alpha_3 & 0 & -\alpha_1 \\ -\alpha_2 & \alpha_1 & 0 \end{pmatrix}$$

we see that the no-twist condition forces $\alpha_3 = 0$ and the no-slip condition forces $\alpha_1 = -\dot{y}$ and $\alpha_2 = \dot{x}$.

For each $(A, q) \in P$ denote by $E_{(A,q)} \subset T_{(A,q)}P$ the two-dimensional subspace consisting of vectors $(\dot{A}, \dot{q})$ such that

$$\dot{q} = (a, b), \quad \dot{A} = \begin{pmatrix} 0 & 0 & a \\ 0 & 0 & b \\ -a & -b & 0 \end{pmatrix} A.$$

Note that the distribution of planes E is smooth and projects onto $\mathbb{R}^2$ under $d\pi_{(A,q)}$ at all points $(A, q) \in P$. We obtain the following lemma.

LEMMA 4.5.12. Suppose that a path $g(t) = (A(t), u(t))$ in $SE(3)$ describes the rolling of S^2 along $t \mapsto \eta(t) \in \mathbb{R}^2$. Then it is a no-slip and no-twist rolling if and only if the curve $\xi(t) := (A(t), \eta(t)) \in P$ satisfies $\xi'(t) \in E_{(A(t),\eta(t))}$ for each t.

It is now easy to prove the existence and uniqueness theorem for the rolling sphere.

PROPOSITION 4.5.13. Let $t \mapsto \eta(t)$ be a C^1 path in $\mathbb{R}^2$ and $g_0 \in SE(3)$ an isometry of $\mathbb{R}^3$ such that $g_0 S^2$ is the unit sphere that touches the plane tangentially at $\eta(0)$. Then there exists a unique curve $t \mapsto g(t)$ in $SE(t)$ such that $g(t) = g_0$ for which $g(t)S^2$ describes the rolling without slipping and twisting of the sphere along η.

PROOF. We write $g_0 = (A_0, u_0)$. By the lemma, we need to find a curve $\xi(t) = (A(t), \eta(t))$ in P that starts at $(A_0, \eta(0))$ such that $\xi'(t)$ for each t is a vector in the plane $E_{\xi(t)}$. Applying the local form of immersions to the curve η, we may assume that there is a local nonzero vector field X on the plane such that $X(\eta(t)) = \eta'(t)$. We can lift X uniquely to a vector field $\overline{X}$ on P defined near $(A_0, \eta(0))$ by setting $\overline{X}(\xi)$ to be the unique vector in E_ξ that projects to $X(\pi(\xi))$ under $d\pi_\xi$. The path in P we want (note that the parametrization is not important) must be the integral curve to $\overline{X}$ that starts at $\xi_0 = (A_0, \eta(0))$. We can now apply the existence and uniqueness theorem for ordinary differential equations. □

6. The Complex Projective Space

CHAPTER 5

Brownian Motion

Let $(\Omega, \mathbb{F}, P)$ be a probability space. A *stochastic process* is a parametrized family of random variables $\{X_t : t \in T\}$ on Ω. It will be assumed for now that X_t takes values in $\mathbb{R}^n$ for each t. The set T will always be here an interval in $\mathbb{R}$, typically $[0, \infty)$, and will be regarded as time parameter of the process. It is possible, however, to do part of the theory that we are going to develop with more general parameter sets, such as subsets of $\mathbb{R}^k$ or a manifold. In this case t might have a spatial interpretation, and X_t would be interpreted as some sort of "random field."

The main process we want to study is Brownian motion, which we will denote by $\{B_t\}$. Our informal discussion of Brownian motion (the velocity process we considered at the beginning of the notes) suggested the following properties that the process B_t is expected to have.

(1) It should not be possible to predict the increments $\Delta B_t := B_{t+\Delta t} - B_t$ from our knowledge of the past of the process. More precisely, ΔB_t, for any positive Δt, is independent of the σ-algebra $\mathcal{B}_t$ generated by $\{B_s : s \leq t\}$. (This is the smallest σ-algebra in $\mathcal{F}$ for which B_s is measurable for each $s \leq t$.) Moreover, the expectation of ΔB_t should be 0 since B_t does not have a preferred direction to go.
(2) The process $\{B_t\}$ should be *stationary*, that is, the probability distribution of ΔB_t (for a fixed Δt) does not depend on t.
(3) We expect the typical Brownian path to be a continuous function. In other words, $t \mapsto B_t(\omega)$ should be continuous for all $\omega \in \Omega$ outside of a set of zero probability.

As suggested by that same informal discussion, if such a process exists, then due to the central limit theorem we expect ΔB_t to be normally distributed with mean 0 and variance $c\Delta t$, where c is some constant, which we suppose for simplicity to be 1. This is because ΔB_t, which was in that discussioin the change of velocity of a Brownian particle, resulted from the contribution of a large number of small momentum transfers that we regarded to be independent and identically distributed. This can be made precise and general, so that if a process $X_t : \Omega \to \mathbb{R}$ satisfies the three properties enumerated above, then the increments ΔX_t are normally distributed random variables with mean 0 and variance Δt.

This claim is proposition 12.4 in *Probability*, Leo Breiman, Classics in Applied Mathematics, 7, SIAM, 1992.

Exercise 5.0.1. Let $\rho(s, x)$ be the gaussian density with mean 0 and variance s, that is,

$$\rho(s, x) = (2\pi s)^{n/2} e^{-\frac{|x|^2}{2s}}.$$

(1) From the general properties of conditional expectation and the above discussion, show that for any integrable $f : \mathbb{R}^n \to \mathbb{R}$,

$$E[f(B_{t+s})|\mathcal{B}_t] = E[f(B_t + \Delta B_t)|\mathcal{B}_t] = \int_{\mathbb{R}^n} f(x + B_t)\rho(s, x)dx.$$

(2) Show that $\rho(t, x)$ satisfies the differential equation:

$$\frac{\partial \rho}{\partial t} = \frac{1}{2}\Delta\rho.$$

(3) If f is a twice differentiable function on $\mathbb{R}^n$ with bounded derivatives show that for every bounded $\mathcal{B}_t$-measurable function $g : \Omega \to \mathbb{R}$,

$$\frac{d}{ds}\langle f \circ B_{t+s}, g\rangle = \langle \frac{1}{2}(\Delta f) \circ B_{t+s}, g\rangle,$$

in which we use the notation $\langle f, g\rangle = E[fg]$. (This requires integration by parts.)

EXERCISE 5.0.2. Show that B_t, viewed as a curve in the Hilbert space $L^2(\Omega, \mathcal{F}, P)$, has the following remarkable property: for any t_0, t_1, t_2, the points $B_{t_0}, B_{t_1}, B_{t_2}$ lie on the vertices of a right triangle.

1. Gaussian Families

put gaussian families back into earlier discussion of probability theory. Rework that section

A real random variable X is called normal if its probability distribution admits the density

$$\rho(x) = \frac{1}{(2\pi)^{1/2}} e^{-\frac{1}{2}x^2}$$

with respect to the Lebesgue measure on the line. Y is called a *gaussian* random variable if $Y = aX + b$, where X is normal and a, b are real constants.

It can be shown that X is normal if and only if $E[e^{iuX}] = e^{-\frac{1}{2}u^2}$ for every $u \in \mathbb{R}$. Y is gaussian if and only if

$$E[e^{iuY}] = e^{-\frac{1}{2}\sigma^2 u^2 + iau},$$

where $a = E[Y]$ and $\sigma^2 = E[(X - a)^2]$.

A random vector $X = (X_1, \dots, X_n)$ taking values in $\mathbb{R}^n$ will be called a *gaussian random vector* (or a gaussian random variable taking values in $\mathbb{R}^n$) if for every linear map $T : \mathbb{R}^n \to \mathbb{R}$ the random variable $T \circ X$ is gaussian. In particular each X_i is gaussian, and X is gaussian if and only if $\lambda_1 X_1 + \cdots + \lambda_n X_n$ is gaussian for all constants $\lambda_1, \dots, \lambda_n$. (We regard a point mass also as a gaussian distribution.)

Define the vector $a = E[X]$ and the positive symmetric matrix $V = (v_{ij})$,

$$v_{ij} = E[(X_i - a_i)(X_j - a_j)].$$

The latter is called the *covariant matrix of* X. If X is a gaussian random vector, its probability distribution is completely determined by the *characteristic function*

$$\phi_X(u) := E[e^{i\langle u, X\rangle}] = e^{i\langle u, a\rangle - \frac{1}{2}u^t V u}$$

for all $u \in \mathbb{R}^n$. In fact, $\phi_X(u)$ is nothing but the Fourier transform (or inverse transform, depending on your sign preference) of the probability density of X:

$$E[e^{i\langle u, X\rangle}] = \int_\Omega e^{i\langle u, X\rangle} dP = \int_{\mathbb{R}^n} e^{i\langle u, x\rangle} d(X_* P)(x) = \int_{\mathbb{R}^n} e^{i\langle u, x\rangle}\rho(x)dx.$$

Let X be a random variable into $\mathbb{R}$ and $Y = (Y_1, \dots, Y_n)$ a random vector. We know that $E[X|Y]$ can be interpreted as the best approximation of X (in the L^2-norm) among functions that are measurable with respect to the σ-algebra generated by Y. In other words, $E[X|Y]$ is the the function of the form $f(Y)$ that minimizes $E[|X - f(Y)|^2]$. Gaussian random variables have the peculiar (and enormously simplifying) property that the minimum is achieved when f is an affine function of Y: $f(Y) = \langle a, Y\rangle + b$. This claim is contained in the next proposition.

But first we observe the following general fact.

LEMMA 5.1.1. Two random variables X, Y from a probability space $(\Omega, \mathcal{F}, P)$ into $\mathbb{R}$ are independent if and only if $E[e^{iuX}e^{ivY}] = E[e^{iuX}]E[e^{ivY}]$. Furthermore, $Y \in L^2(\Omega, \mathcal{F}, P)$ is a centered gaussian random variable if and only if

$$\phi_Y(u) := E[e^{iuY}] = e^{-\frac{1}{2}u^2 E[Y^2]}.$$

PROOF. We leave the details to the reader and only make the following observation. The first claim is equivalent to the statement that for all measurable functions f, g, $E[f(X)g(Y)] = E[f(X)]E[g(Y)]$. (That independence implies the stated property of the characteristic functions is immediate.) Let $\hat{f}$ and $\hat{g}$ denote the Fourier transforms of f and g, respectively. Therefore we have

$$f(X) = \int_{\mathbb{R}^n} \hat{f}(u)e^{iuX}du$$

with a similar formula for $g(Y)$. Therefore

$$\begin{aligned}
E[f(X)g(Y)] &= E\left[\int_{\mathbb{R}^n} \hat{f}(u)e^{iuX}du \int_{\mathbb{R}^n} \hat{g}(v)e^{ivY}dv\right] \\
&= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \hat{f}(u)\hat{g}(v)E[e^{iuX}e^{ivY}]dudv \\
&= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \hat{f}(u)\hat{g}(v)E[e^{iuX}]E[e^{ivY}]dudv \\
&= E\left[\int_{\mathbb{R}^n} \hat{f}(u)e^{iuX}du\right] E\left[\int_{\mathbb{R}^n} \hat{g}(v)]e^{ivY}\right] dv.
\end{aligned}$$

□

A linear subspace $H \subset L^2(\Omega, \mathcal{F}, P)$ (of functions into $\mathbb{R}$) will be called a *gaussian family* if every $X \in H$ is a centered (i.e., $E[X] = 0$) gaussian random variable.

PROPOSITION 5.1.2. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $H \subset L^2(\Omega, \mathbb{F}, P)$ a gaussian family. Then

(1) The closure of H is also a gaussian family.
(2) If $H' \subset L^2(\Omega, \mathcal{F}, P)$ is a second gaussian family and $\mathcal{H}$ and $\mathcal{H}'$ (the respective σ-algebras) are independent, then $H + H'$ (an orthogonal direct sum) is again a gaussian family.
(3) Let K be a non-empty subset of H, $K^\perp \cap H$ the orthogonal complement of K in H, and denote by $\mathcal{K}$ and $\mathcal{K}^\perp$ the σ-algebras generated by K and $K^\perp \cap H$, respectively. Then $\mathcal{K}$ and $\mathcal{K}^\perp$ are independent.

PROOF. For part (1), let Y_n be a sequence in H that converges to $Y \in L^2(\Omega, \mathcal{F}, P)$. Due to the previous lemma, to see that Y is also a centered gaussian random variable it suffices to note that

$$\begin{aligned}E[e^{iuY}] &= \lim_{n\to\infty} E[e^{iuY_n}]\\ &= \lim_{n\to\infty} e^{-\frac{1}{2}u^2 E[Y_n^2]}\\ &= e^{-\frac{1}{2}u^2 E[Y^2]}.\end{aligned}$$

The lemma also yields part (2). In fact, since H and H' are independent, they are also orthogonal subspaces, and all that is needed is to check that any sum $X+Y$, $X \in H$, $Y \in H'$, is also a centered gaussian random variable. But

$$\begin{aligned}E[e^{iu(X+Y)}] &= E[e^{iuX}]E[e^{iuY}]\\ &= e^{-\frac{1}{2}u^2 E[X^2]} e^{-\frac{1}{2}u^2 E[Y^2]}\\ &= e^{-\frac{1}{2}u^2 (E[X^2]+E[Y^2])}\\ &= e^{-\frac{1}{2}u^2 E[(X+Y)^2]},\end{aligned}$$

where in the last step we used that $E(XY) = 0$.

To prove part (3), it suffices to show that for every finite subset $\{X_1, \dots, X_l\}$ of K and $\{Y_1, \dots, Y_k\}$ of $K^\perp \cap H$, and for every $a_1, \dots, a_l, b_1, \dots, b_k$, the centered gaussian random variables $X = a_1 X_1 + \cdots + a_l X_l$ and $Y = b_1 Y_1 + \cdots + b_k Y_k$ satisfy

$$E[e^{iX} e^{iY}] = E[e^{iX}]E[e^{iY}].$$

But $X+Y$ is centered gaussian, so

$$\begin{aligned}E[e^{iX} e^{iY}] &= E[e^{i(X+Y)}]\\ &= e^{-\frac{1}{2}E[(X+Y)^2]}\\ &= e^{-\frac{1}{2}(E[X^2]+2E[XY]+E[Y^2]}\\ &= e^{-\frac{1}{2}(E[X^2]+E[Y^2])}\\ &= e^{-\frac{1}{2}E[X^2]} e^{\frac{1}{2}E[Y^2]}\\ &= E[e^{iX}]E[e^{iY}].\end{aligned}$$

□

2. Lévy's Construction of Brownian Motion

Let $\mathcal{P}^0(\mathbb{R}^n)$ be the space of all continuous paths $\gamma : [0, \infty) \to \mathbb{R}^n$ such that $\gamma(0) = 0$. With the topology of uniform convergence on compact intervals, this is a separable topological space that admits a complete metric. We fix a probability space $(\Omega, \mathcal{F}, P)$.

In this section we obtain Brownian motion by constructing a random variable $B : \Omega \to \mathcal{P}^0(\mathbb{R}^n)$ having the following properties. Let $B_t = \pi_t \circ B$, where $\pi_t : \mathcal{P}^0(\mathbb{R}^n) \to \mathbb{R}^n$ is the evaluation of B at time t. Then for every t and Δt, $\Delta B_t := B_{t+\Delta t} - B_t$ will be a centered gaussian random variable of variance $\sigma^2 \Delta t$ (we will take $\sigma = 1$), which is independent of $\mathcal{B}_t$ (the σ-algebra generated by the B_s for $s \le t$).

For concreteness, and since no generality is lost, the probability space $(\Omega, \mathcal{F}, P)$ will be taken to be $[0, 1]$, with the σ-algebra generated by intervals, and P the Lebesgue measure.

In order to know what we should be looking for, suppose for the moment that B exists and let us see what it gives. First, $B_t - B_s$ should be a centered gaussian random variable, and if (s,t) and (s',t') are disjoint intervals then $B_{t'} - B_{s'}$ and $B_t - B_s$ are independent. So the collection of increments $\{B_t - B_s : 0 \leq s < t\}$ should be a gaussian family. At a minimum, our probability space must contain an infinite collection of independent centered gaussian random variables. (It turns out that countably many independent normal random variables will suffice.) Anticipating this need, we recall the following lemma, which was established in an earlier section.

Lemma 5.2.1. On $[0,1]$ (with the Borel σ-algebra and the Lebesgue mesure) there exists a countable family $N_1, N_2, \ldots$ of independent normal random variables.

It will be convenient, for reasons that will become clear in a moment, to index the N_i using non-negative dyadic rationals $k/2^n$ (in whatever order we wish). All dyadic rationals, with the exception of the even integers, can be expressed uniquely as $(2m+1)/2^n$ for m, n non-negative integers. (Note that $(2m+1)/2^n = (2l+1)/2^k$ if and only if $m = l$ and $n = k$.) We will write the family of random variables granted by the lemma as $\{Y_m, Y_{\frac{2m+1}{2^n}}\}$ where m ranges over the non-negative integers and n over the positive (non-zero) integers.

The approach now will be to use this family of normal random variables to construct a sequence $B^{(0)}, B^{(1)}, B^{(2)}, \ldots$ of piecewise linear random processes (random flights) which will converge to a process B having the desired properties.

Define $B^{(0)}$ so that $B_0^{(0)} = 0$ (i.e., $\omega \mapsto B_0^{(0)}(\omega)$ is constant equal to 0, for $\omega \in \Omega = [0,1]$) and, for $t > 0$,

$$B_t^{(0)} = B_m^{(0)} + (t-m)Y_m, \text{ for } m \in \mathbb{N} \text{ and } t \in [m, m+1].$$

(If B already existed and we wished to find a polygonal approximation $B_t^{(0)}$ such that $B_m^{(0)} = B_m$, then $B_t^{(0)} = B_m + (t-m)L_m$, where $L_m = B_{m+1} - B_m$.)

The other $B^{(k)}$ will be defined inductively. Define first

$$\lambda_n(t) = \begin{cases} 2^n(t - \frac{m}{2^{n-1}}) & \text{if } \frac{m}{2^{n-1}} \leq t \leq \frac{m+1/2}{2^{n-1}}, \\ 2 - 2^n(t - \frac{m}{2^{n-1}}) & \text{if } \frac{m+1/2}{2^{n-1}} \leq t \leq \frac{m+1}{2^{n-1}}. \end{cases}$$

Now define

$$B_t^{(n)} = B_t^{(n-1)} + 2^{-(n+1)/2}\lambda_n(t)Y_{\frac{2m+1}{2^n}} \text{ for } \frac{m}{2^{n-1}} \leq t \leq \frac{m+1}{2^{n-1}}.$$

Before continuing, let us take a little time to understand where this definition is coming from. Suppose again that B is already given and that we are looking for the polygonal approximation $B_t^{(n)}$ of B_t that coincides with B_t when t is a dyadic rational of the form $(2m+1)/2^n$ (n fixed). Write $\Delta_{m,n}B := B_{\frac{m+1}{2^n}} - B_{\frac{m}{2^n}}$ and

$D_{mn} = B_{\frac{m+1/2}{2^{n-1}}} - B^{(n-1)}_{\frac{m+1/2}{2^{n-1}}}$. Observe that

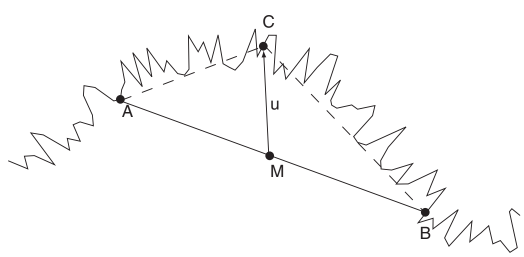
$$\begin{aligned} D_{mn} &= B_{\frac{m+1/2}{2^{n-1}}} - B^{(n-1)}_{\frac{m+1/2}{2^{n-1}}} \\ &= B_{\frac{m+1/2}{2^{n-1}}} - B^{(n-1)}_{\frac{m+1/2}{2^{n-1}}} + \frac{1}{2}\left(2B^{(n-1)}_{\frac{m+1/2}{2^{n-1}}} - (B^{(n-1)}_{\frac{m}{2^{n-1}}} + B^{(n-1)}_{\frac{m+1}{2^{n-1}}})\right) \\ &= B_{\frac{m+1/2}{2^{n-1}}} - \frac{1}{2}\left(B_{\frac{m}{2^{n-1}}} + B_{\frac{m+1}{2^{n-1}}}\right) \\ &= \frac{1}{2}\left(2B_{\frac{2m+1}{2^n}} - B_{\frac{2m}{2^n}} - B_{\frac{2m+2}{2^n}}\right) \\ &= \frac{1}{2}\left(\Delta_{2m,n}B - \Delta_{2m+1,n}B\right). \end{aligned}$$

Therefore the D_{mn} (for a fixed n) would be independent centered gaussian random variables with variance $1/2^{n+1}$. It should be a simple exercise to check that $B^{(n)}$ would take the form

$$B^{(n)}_t = B^{(n-1)}_t + 2^{-(n+1)/2}\lambda_n(t)[2^{(n+1)/2}D_{mn}],$$

where the term between square brackets is normal. It is this term that we replace with the normal random variables $Y_{\frac{2m+1}{2^n}}$ in the defintion of $B^{(n)}$.

In the picture shown below, the rough path represents a sample path of B, the solid line represents a piece of B^{n-1} and the broken line a piece of B^n. The points are: $A = B^{(n-1)}_{m/2^{n-1}}$, $B = B^{(n-1)}_{(m+1)/2^{n-1}}$, M is the middle point between A and B, so it is $B^{(n-1)}_{(m+1/2)/2^{n-1}}$ and $C = B_{(m+1/2)/2^{n-1}}$. The vector u is D_{mn}.



We can now return to the proof of existence of B. We wish to show that the sequence $B^{(n)}$ converges to a process B having the properties of Brownian motion.

For each fixed n

$$B^{(n)}_{\frac{m+1}{2^n}} - B^{(n)}_{\frac{m}{2^n}}$$

is a sequence of mutually independent centered gaussian random variables with variance 2^{-n}. If we can show that $B^{(n)}_t$ almost surely converges (uniformly in t on compact intervals of $[0,\infty)$) to a limit B, then B will be the process we want.

Note: for $m/2^n \leq t \leq (m+1)/2^n$

$$|B^{(n)}_t - B^{(n-1)}_t| \leq 2^{-\frac{n+1}{2}}|Y_{\frac{2m+1}{2^n}}|.$$

Therefore, for any fixed $T \in (0, \infty)$,

$$\sup_{s\in[0,T]} |B_s^{(n)} - B_s^{(n-1)}| \leq 2^{-\frac{n+1}{2}} \max_{m \leq 2^{n-1}T} |Y_{\frac{2m+1}{2^n}}|$$

$$\leq 2^{-\frac{n+1}{2}} \left(\sum_{m\leq 2^{n-1}T} |Y_{\frac{2m+1}{2^n}}|^4 \right)^{\frac{1}{4}}.$$

Applying Jensen's inequality (for the concave function $\phi(x) = x^{1/4}$) we arrive at

$$E[\sup_{s\in[0,T]} |B_s^{(n)} - B_s^{(n-1)}|] \leq 2^{-\frac{n+1}{2}} \left(\sum_{m\leq 2^{n-1}T} E[Y^4_{\frac{2m+1}{2^n}}] \right)^{\frac{1}{4}}.$$

Writing $C = E[Y^4_{\frac{2m+1}{2^n}}]$ (so that C is the value $\int_{\mathbb{R}} x^4 \rho(x) dx$, for the standard gaussian density ρ), we have

$$E[\sup_{s\in[0,T]} |B_s^{(n)} - B_s^{(n-1)}|] \leq (2TC)^{1/4} 2^{-\frac{n}{4}}.$$

Therefore

$$\sum_{n=1}^{\infty} E[\sup_{s\in[0,T]} |B_s^{(n)} - B_s^{(n-1)}|] < \infty.$$

It follows that there exists a subset $\Omega' \subset \Omega$ of probability 1 such that, for each $\omega \in \Omega'$, the sequence $t \in [0,T] \mapsto B_t^{(n)}(\omega)$ converges to a continuous function $t \mapsto B_t(\omega)$. The result is a measurable process B_t with the desired properties. expand on this last step.

To obtain the n-dimensional version of B, simply produce as above n independent processes, B_i, and form the random vector $B = (B_1, \ldots, B_n)$.

3. The Wiener Measure

Recall that $\mathcal{P}^0(\mathbb{R}^n)$ denotes the space of continuous paths $\gamma : [0, \infty) \to \mathbb{R}^n$ such that $\gamma(0) = 0$. For simplicity of notation, we take $n = 1$, and only consider the general case when it does not follow from an obvious modification of the one-dimensional situation.

Let $B : \Omega \to \mathcal{P}^0(\mathbb{R})$ be the Brownian motion process constructed in the previous section. Let $\mu = B_* P$ be the probability distribution of B, a probability measure on $\mathcal{P}^0(\mathbb{R})$. Instead of carrying around the probability space $(\Omega, \mathcal{F}, P)$ of the previous section it will be more convenient to work only with $\Omega^0 := \mathcal{P}^0(\mathbb{R})$ and the probability μ on the Borel σ-algebra, $\mathcal{B}$, of Ω^0. The upper index 0 indicates that the process starts at the origin at $t = 0$. The random variable B_t will then be regarded as a map from Ω^0 to $\mathbb{R}$, which is simply the evaluation of a path at time t: $B_t(\omega) = \omega(t)$. We continue to use the notation

$$\rho(t, x) := \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2t}x^2}.$$

Proposition 5.3.1. *The measure μ is the only probability measure on $(\Omega^0, \mathcal{B})$ for which either of the following equivalent properties holds:*

(1) *For each $t \in [0, \infty)$ and $s > 0$, the increment*

$$\Delta B_t : \omega \mapsto B_{t+s}(\omega) - B_t(\omega)$$

is a centered gaussian random variable with variance s which is independent of the σ-algebra $\mathcal{B}_t$ generated by $\{B_r : r \leq t\}$.

(2) *For every bounded C^∞ function f on $\mathbb{R}^n$, for all $\mathcal{B}_t$-measurable set $A \subset \Omega^0$, and all $s > 0$,*

$$E\left[\left(f \circ B_{t+s} - f \circ B_t - \int_t^{t+s} \frac{1}{2}(\Delta f) \circ B_s ds\right) \chi_A\right] = 0.$$

PROOF. We only give a sketch the proof. For any positive integer k, reals $0 = t_0 < t_1 < \cdots < t_k$, and measurable subsets $A_1, \ldots, A_k$ of $\mathbb{R}$, define the set

$$C_i = \{\omega \in \Omega^0 : B_{t_i}(\omega) - B_{t_{i-1}}(\omega) \in A_i\}$$

and $C = C_1 \cap \cdots \cap C_k$. Since the increments of B are independent, we must have

$$\mu(C) = \mu(C_1) \ldots \mu(C_k) = \prod_{i=1}^{k} \int_{A_i} \rho(t_i - t_{i-1}, x) dx.$$

This formula shows that μ is uniquely specified on a σ-algebra containing $\mathcal{B}_t$ for all $t \geq 0$. But it can be shown that this σ-algebra coincides with $\mathcal{B}$, which proves the first claim.

That (1) implies (2) follows from a previous exercise (this is essentially integration by parts). To see the converse, suppose that the property stated in (2) holds for some measure ν. We want to check that $\nu = \mu$.

This can be done by applying the identity in (2), now holding for ν, to the function $f(x) = e^{i\langle u,x\rangle}$ (so as to get the characteristic function) and showing that $B_{t+s} - B_t$ is also centered gaussian with covariance sI (where I is the identity matrix). □

elaborate

4. The Martingale Interpretation

Part (2) of the previous proposition can be restated as follows. The Wiener measure μ is characterized by the property that, given $\mathcal{B}_t$, the conditional expectations of $f \circ B_{t+s} - f \circ B_t$ and $\int_t^{t+s} \frac{1}{2}[\Delta f] \circ B_s ds$ coincide, for any bounded infinitely differential function f on $\mathbb{R}^n$. This can be rephrased by saying that

$$M_f^\Delta(t) = f \circ B_t - \int_0^t \frac{1}{2}[\Delta f] \circ B_s ds$$

is a "conditionally constant" function of t relative to the family $\mathcal{B}_t$, that is,

$$E[M_f^\Delta(t+s)|\mathcal{B}_t] = M_f^\Delta(t)$$

almost surely in μ. But a process which is conditionally constant is precisely what is meant by a *martingale*.

More formally, given a probability space $(\Omega, \mathcal{F}, P)$ and a non-decreasing family $\mathcal{F}_t$ of sub-σ-algebras of $\mathcal{F}$, then a process X_t is a martingale on Ω if the following hold:

(1) The map $X : [0, \infty) \times \Omega \to \mathbb{R}$ is *adapted* to $\{\mathcal{F}_t\}$, that is, for each $t \in [0, \infty)$, $X(t, \cdot)$ is a $\mathcal{F}_t$-measurable and integrable;
(2) for each $\omega \in \Omega$, $t \mapsto X(t, \omega)$ is right-continuous;
(3) it is conditionally constant:

$$X_{t_1} = E[X_{t_2}|\mathcal{F}_{t_1}]$$

for all $0 \leq t_1 < t_2$.

5. The Martingale Problem for Hörmander Operators

We have seen that the Wiener measure can be characterized in terms of the Laplace operator Δ. More generally, given a manifold M and a differential operator L on the space $C^\infty(M)$ of real valued infinitely differentiable functions on M, one can construct in a similar way a probability measure on the space of continuous paths in M.

In other words, the general problem is to construct a probability measure P on the space $\mathcal{P}(M)$ of all continuous paths $\gamma : [0, \infty) \to M$ so that P is supported on the set of paths starting at p_0 (i.e., $\gamma(0) = p_0$) and for some sufficiently big class of smooth functions on M

$$M_f^L(t) := f \circ B_t - \int_0^t [Lf] \circ B_s ds$$

is a P-martingale relative to the σ-algebra generated by the B_s, $s \leq t$. P will then be said to solve the *martingale problem* for L starting at p_0. Notice that we are using the notation $B_t : \mathcal{P}(M) \to M$, $B_t(\gamma) = \gamma(t)$, as for $\mathbb{R}^n$. Of course, for a general manifold, it no longer makes sense to write expressions like $B_t - B_s$.

Exercise 5.5.1. Let $F : M \to N$ be a diffeomorphism between two smooth manifolds. Also denote by F the induced map from $\mathcal{P}(M)$ to $\mathcal{P}(N)$ defined by $\gamma \mapsto F \circ \gamma$. Let L be a differential operator on M and suppose that P is a probability on $\mathcal{P}(M)$ that solves the martingale problem for L with initial point $p \in M$. Then $F_* P$ solves the martingale problem for $F_* L$ with initial point $F(p)$.

a. $L = X$, a vector field. Let X be a smooth vector field on M. We suppose that X is *complete*, that is, its flow Φ_t is defined for all $t \in \mathbb{R}$. This is the case, for example, if M is compact.

Proposition 5.5.2. A probability P on $\mathcal{P}(M)$ solves the martingale problem for $L = X$ if and only if P is the unit point mass supported on $t \mapsto \gamma(t) = \Phi_t(p_0)$.

Proof. Recall that Φ_t satisfies the differential equation

$$\frac{d\Phi_t}{dt}(p_0) = X(\Phi_t(p_0)).$$

This implies that for every C^1 real valued fucntion $f : M \to \mathbb{R}$,

$$\frac{d}{dt}(f(\Phi_t(p_0))) = (Xf)(\Phi_t(p_0)),$$

which can be integrated in t to yield

$$f(\Phi_{t_2}(p_0)) - f(\Phi_{t_1}(p_0)) = \int_{t_1}^{t_2} (Xf)(\Phi_s(p_0)) ds. \tag{5.1}$$

Let P be the probability on $\mathcal{P}(M)$ given by the unit point mass supported on $t \mapsto \gamma(t) = \Phi_t(p_0)$. We claim that P solves the martingale problem for L. In fact, since

$$\begin{aligned} M_f^L(t_2) - M_f^L(t_1) &= f \circ B_{t_2} - f \circ B_{t_1} - \left(\int_0^{t_2} (Xf) \circ B_s ds - \int_0^{t_1} (Xf) \circ B_s ds \right) \\ &= f \circ B_{t_2} - f \circ B_{t_1} - \int_{t_1}^{t_2} (Xf) \circ B_s ds, \end{aligned}$$

then Equation 5.1 implies that $M_f^L(t_2) - M_f^L(t_1) = 0$ almost surely for P. In particular, $M_f^L(t)$ is (conditionally) constant.

For the converse, suppose that P solves the martingale problem for L and $B_0 = p_0$ P-almost surely. Let f be any C^∞ function on M with compact support and set $u(t, p_0) := f(\Phi_t(p_0))$. Now take the expectation of $M_f^L(t_2) - M_f^L(t_1)$:

$$
\begin{aligned}
0 &= E^P[M_f^L(t_2) - M_f^L(t_1)] \\
&= E^P[f \circ B_{t_2}] - E^P[f \circ B_{t_1}] - E^P[\int_{t_1}^{t_2} (Xf) \circ B_s ds] \\
&= E^P[f \circ B_{t_2}] - E^P[f \circ B_{t_1}] - \int_{t_1}^{t_2} E^P[(Xf) \circ B_s] ds.
\end{aligned}
$$

This shows that $E^P[f \circ B_t]$ is C^1 and

$$\frac{d}{dt} E^P[f \circ B_t] = E^P[(Xf) \circ B_t].$$

If instead of f we apply the previous equation to $f \circ \Phi_s$ then, as $f(\Phi_s(B_t)) = u(s, B_t)$, we obtain

$$\frac{d}{dt} E^P[u(s, B_t)] = E^P[(Xu)(s, B_t)].$$

We now have, for a fixed T:

$$
\begin{aligned}
E^P[f \circ B_T] - f(\Phi_T(p_0)) &= E^P[u(0, B_T)] - E^P[u(T, B_0)] \\
&= \int_0^T \frac{d}{ds} E^P[u(T-s, B_s)] ds \\
&= \int_0^T \left(\frac{\partial}{\partial t_1}|_{t_1 = T-s} E^P[u(t_1, B_{t_2})] \frac{dt_1}{ds} + \frac{\partial}{\partial t_2}|_{t_2 = s} E^P[u(t_1, B_{t_2})] \frac{dt_2}{ds} \right) ds \\
&= \int_0^T \left(-E^P[\frac{\partial}{\partial t_1}|_{t_1 = T-s} u(t_1, B_s)] + \frac{\partial}{\partial t_2}|_{t_2 = s} E^P[u(T-s, B_{t_2})] \right) ds \\
&= \int_0^T \left(-E^P[Xu(T-s, B_s)] + E^P[Xu(T-s, B_s)] \right) ds \\
&= 0.
\end{aligned}
$$

But $E^P[f \circ B_T] = f(\Phi_T(p_0))$ for all f is precisely what it means for P to be concentrated on $t \mapsto \Phi_t(p_0)$. □

b. Sums of Squares of Commuting Vector Fields. Let $X_1, X_2, \ldots, X_k$ be complete vector fields on a manifold M and define $L = X_1^2 + \cdots + X_k^2$. The next proposition will be needed for solving the martingale problem for L. The partial derivative with respect to t_i will be denoted D_i.

Proposition 5.5.3. Fix $p_0 \in M$ and denote by $\Phi_t^{(i)}$ the flow of X_i. Define the map

$$\Phi : t = (t_1, \ldots, t_k) \in \mathbb{R}^k \mapsto (\Phi_{t_1}^{(1)} \circ \cdots \circ \Phi_{t_k}^{(k)})(p_0) \in M.$$

Then for every C^∞ function h on M and every positive integer l we have

$$D_{i_1} \ldots D_{i_l}(h \circ \Phi) = (X_{i_1} \ldots X_{i_l} h) \circ \Phi.$$

In particular,

$$\Delta(h \circ \Phi) = (Lh) \circ \Phi$$

where Δ is the euclidian Laplacian on $\mathbb{R}^k$.

PROOF. Once we show that $D_i(h \circ \Phi) = (X_i h) \circ \Phi$ the general case follows by induction. Let $t = (t_1, \ldots, t_k)$. Then

$$\begin{aligned} D_i(h \circ \Phi)_t &= \frac{\partial}{\partial t_i}(h \circ \Phi^{(1)}_{t_1} \circ \cdots \circ \Phi^{(k)}_{t_k})(p_0) \\ &= \frac{\partial}{\partial t_i}(h \circ \Phi^{(i)}_{t_i} \circ \Phi^{(1)}_{t_1} \circ \cdots \circ \Phi^{(i-1)}_{t_{i-1}} \circ \Phi^{(i+1)}_{t_{i+1}} \circ \cdots \circ \Phi^{(k)}_{t_k})(p_0) \\ &= (Xh)(\Phi^{(i)}_{t_i} \circ \Phi^{(1)}_{t_1} \circ \cdots \circ \Phi^{(i-1)}_{t_{i-1}} \circ \Phi^{(i+1)}_{t_{i+1}} \circ \cdots \circ \Phi^{(k)}_{t_k})(p_0) \\ &= (Xh)(\Phi^{(1)}_{t_1} \circ \cdots \circ \Phi^{(k)}_{t_k})(p_0) \\ &= (Xh)(\Phi(t)). \end{aligned}$$

□

It will be convenient to use at different moments the different notations

$$\Phi(x) = \Phi_{p_0}(x) = \Phi_x(p_0),$$

where $x \in \mathbb{R}^k$ and $p_0 \in M$. This allows us to write

$$\begin{aligned} \Phi_{p_0} \circ B_{t+\tau} &= \Phi_{B_{t+\tau}}(p_0) \\ &= \Phi_{B_{t+\tau}-B_t}(\Phi_{B_t}(p_0)) \\ &= \Phi_{\Phi_{B_t}(p_0)} \circ (B_{t+\tau} - B_t). \end{aligned}$$

Since $B_{t+\tau} - B_t$ is independent of the σ-algebra $\mathcal{B}_t$ (generated by the B_s, $s \leq t$) we can write (see Exercisekey), for every C^∞ function $f : M \to \mathbb{R}$ with compact support,

$$E[f \circ \Phi_{p_0} \circ B_{t+\tau} | \mathcal{B}_t] = \int_{\mathbb{R}^k} f(\Phi_{\Phi_{B_t}(p_0)}(x))\rho(\tau, x)dx. \tag{5.2}$$

Define a map from $\mathcal{P}(\mathbb{R}^k)$ to $\mathcal{P}(M)$ by $\gamma \mapsto \Phi_{p_0} \circ \gamma$. Calling this map also Φ, it makes sense to consider the push-forward of the Wiener measure under it.

PROPOSITION 5.5.4. If P_0 is the Wiener probability measure on $\mathcal{P}(\mathbb{R}^k)$ (for the initial point 0), then $P := \Phi_* P_0$ solves the martingale problem for L (for the initial point p_0). In other words, if γ is a typical Brownian path in $\mathbb{R}^k$, then $\Phi \circ \gamma$ is a typical path associated to L. Furthermore, P is the unique solution to the martingale problem for L.

PROOF. We first show that P solves the martingale problem for L. Let f be any C^∞ function on M with compact support. The claim is, then, that the conditional expectation of

$$f \circ \Phi \circ B_{t_2} - f \circ \Phi \circ B_{t_1} - \int_{t_1}^{t_2} (Lf) \circ \Phi \circ B_s ds$$

given $\mathcal{B}_{t_1}$ (the σ-algebra generated by the B_s, $s \leq t_1$) is zero. Let $A \in \mathcal{B}_{t_1}$ and define

$$F(s) := \int_A f \circ \Phi \circ B_{t_1+s} dP_0 =: E^{P_0}_A[f \circ \Phi \circ B_{t_1+s}].$$

Using Equation 5.2, Exercise 5.0.1, and the fact that Φ intertwines Δ and L, we get:

$$\begin{aligned} F'(s) &= E_A^{P_0}[\Delta(f \circ \Phi) \circ B_{t_1+s}] \\ &= E_A^{P_0}[(Lf) \circ \Phi \circ B_{t_1+s}]. \end{aligned}$$

Integrating in s, from t_1 to t_2, stablishes the claim.

To prove uniqueness, suppose that P is also a solution of the martingale problem for L. It follows from Exercise 5.5.1 that if $\varphi : U \subset \mathbb{R}^n \to M$ is a local parametrization of M and f is a C^∞ function on M with compact support such that the closure of the support is contained in $\varphi(U)$, then $M_f^L(t)$ is P-conditionally constant (a martingale) for a probability P on $\mathcal{P}(\varphi(U))$ if and only if $M_{f\circ\varphi}^{\varphi_*^{-1}L}(t)$ is conditionally constant for $\varphi_*^{-1}P$. Therefore, it suffices to consider the case $M = \mathbb{R}^n$, and thus suppose that the X_i are vector fields on $\mathbb{R}^n$.

Let $\psi : \mathbb{R}^k \to [0,1]$ be a C^∞ function that is identically 1 on the ball of radius 1 and center 0 and that vanishes on the complement of the ball of radius 2. For each $R > 1$ define

$$u_R(t,y) := \psi\left(|y|/R\right) \int_{\mathbb{R}^k} \psi(x/R) f(\Phi_y(x)) \rho(t,x) dx$$

where ρ is the Gaussian density function with mean 0 and variance t. Note that u_R is C^∞ on $(0,\infty) \times \mathbb{R}^k$. An integration by parts gives:

$$\begin{aligned} \frac{\partial u_R}{\partial t}(t,y) &= \psi\left(|y|/R\right) \int_{\mathbb{R}^k} \psi(x/R) f(\Phi_y(x)) \frac{\partial \rho}{\partial t}(t,x) dx \\ &= \psi\left(|y|/R\right) \int_{\mathbb{R}^k} \psi(x/R) f(\Phi_y(x)) \frac{1}{2} (\Delta\rho)(t,x) dx \\ &= \psi\left(|y|/R\right) \int_{\mathbb{R}^k} \frac{1}{2} \Delta(\psi(\cdot/R) f(\Phi_y(\cdot)))|_{\cdot=x} \rho(t,x) dx \\ &= \psi\left(|y|/R\right) \int_{\mathbb{R}^k} \psi(x/R) \frac{1}{2} \Delta(f(\Phi_y(\cdot)))|_{\cdot=x} \rho(t,x) dx + H \\ &= \psi\left(|y|/R\right) \int_{\mathbb{R}^k} \psi(x/R) \frac{1}{2} Lf(\Phi_y(x)) \rho(t,x) dx + H, \end{aligned}$$

where H is equal to

$$\psi\left(|y|/R\right) \int_{\mathbb{R}^k} \frac{1}{2} \frac{1}{R^2} \Delta\psi(x/R) f(\Phi_y(x))) + \langle \frac{1}{R} \nabla\psi(x/R), (\nabla f)(\Phi_y(\cdot))|_{\cdot=x} \rangle \rho(t,x) dx.$$

We have used above the fact that $\Delta(f \circ \Phi_y) = (Lf) \circ \Phi_y$. Note also that $(Lf)(\Phi_y(x)) = L(f \circ \Phi.)|_{\cdot=y}$. Putting all this together we obtain the estimate:

$$\left| \frac{\partial u_R}{\partial t}(t,y) - \frac{1}{2} L u_R(t,y) \right| \leq \frac{C}{R},$$

for some constant C independent of R. Therefore, for all t_1, t_2 such that $0 \leq t_1 < t_2 < \infty$ and a $\mathbb{B}_{t_1}$-measurable set A,

$$\left| \frac{d}{dt} E_A^P [u_R(t_2 - t, B_t] \right| \leq \frac{C}{R}$$

for all t, $t_1 < t < t_2$. Integrating in t, and using that $u_R(0, y) = \psi(y/R)f(y)$, we get:

$$\left|E_A^P[\psi(R^{-1}B_{t_2})f \circ B_{t_2}] - E_A^P[u_R(t_2 - t_1, B_{t_1})]\right| \leq \frac{C(t_2 - t_1)}{R}.$$

Let R go to ∞ we arrive at

$$E_A^P[f \circ B_{t_2}] = E_A^P[u(t_2 - t_1, B_{t_1})],$$

where

$$u(t, y) = \int_{\mathbb{R}^k} f(\Phi_y(x))\rho(t, x)dx.$$

Consequently, for any solution P of the martingale problem for L we have

$$E^P[f \circ B_{t_2}|\mathcal{B}_{t_1}] = u(t_2 - t_1, B_{t_1}).$$

But the last equation uniquely characterize the law of B_t for the probability P. It follows that P is unique. □ elaborate

EXERCISE 5.5.5. Solve the martingale problem for the operator

$$L = X_0 + \frac{1}{2}\sum_{i=1}^{k} X_i^2$$

for commuting vector fields $X_0, \ldots, X_k$.

6. Stochastic Differential Equations

The approach to stochastic differential equations (and stochastic processes on manifolds) that we are going to take can be naturally cast in the language of control theory. Indeed, we will think of a stochastic process on a manifold as the solution of a controlled ordinary differential equation (determined by vector fields on the manifold) for which the control function is a Brownian motion on the space of control parameters.

a. Some Definitions about Controlled Differential Equations. We will consider control systems of the following form. Let U be a subset of $\mathbb{R}^m$ (that will later be taken to be $\mathbb{R}^m$ itself), called the *control set*. The dynamics of the system under consideration is specified by a differentiable map $F : M \times U \to TM$ such that for each $u \in U$, $p \in M \mapsto F_u(p) := F(p, u) \in T_pM$ is a differentiable vector field on M.

Velimir Jurdjevic - Geometric Control Theory, Cambridge Studies in Advanced Mathematics 51, Cambridge University Press, 1997

Given a *control function* $u : [0, \infty) \times M \to U$, one obtains a differential equation on TM given by

$$\frac{dp}{dt} = F(p, u(t, p)). \tag{5.1}$$

The control function $u(t, p)$ is called a *closed-loop* control, or a *feed-back* control, if it only depends on p, and it is called an *open-loop* control if it only depends on t. We will be mainly concerned with control systems of the following form: Let $X_0, X_1, \ldots, X_m$ be vector fields on M and write $u = (u_1, \ldots, u_m)$. We say that Equation 5.1 describes an *affine control system* if

$$F(p, u) = X_0 + u_1X_1 + \cdots + u_mX_m.$$

One is typically interested in finding a control function u that will steer trajectories of the differential equation 5.1 to behave in desired ways. Here, however, we are interested in control functions given by "white noise" as will be explained later. Anyhow, if the control function is a random variable, $u : \Omega \times [0, \infty) \to \mathbb{R}^m$, then the solution will be a random process taking values in M, as will be discussed further on.

b. Example: the Serret-Frenet Equations. Let $\gamma(t)$ be a differentiable curve in $\mathbb{R}^3$, which will be supposed for simplicity to be parametrized by arclength. Therefore $v(t) := \gamma'(t)$ is a unit vector field along γ. One defines the *curvature* function $k(t)$ and the *principal normal* vector field n of γ by the equation $\frac{dv}{dt} = \kappa n$, where n is a unit vector, necessarily orthogonal to v, since $|v(t)| = 1$ for all t. The principal normal is well defined whenever the derivative of $v(t)$ is not zero, which is the case whenever $\gamma(t)$ is "bending away" from its tangent line. Define the *binormal* vector field $b(t)$ so that $v(t), n(t), b(t)$ constitute a positive orthonormal frame. We have therefore a well-defined orthonormal frame (base of $\mathbb{R}^3$) at each point $\gamma(t)$ where the curve is not straight up to second order derivatives.

Proposition 5.6.1 (Serret-Frenet Formulae). Suppose that γ is a (twice) differentiable curve in $\mathbb{R}^3$ parametrized by arclength and that $\gamma''(t)$ does not vanish. Then there are functions $\kappa(t)$ and $\tau(t)$ such that

$$
\begin{aligned}
dv/dt &= \kappa n \\
dn/dt &= -\tau b - \kappa v \\
db/dt &= \tau n.
\end{aligned}
$$

Proof. Let $g(t)$ denote the element of $SO(3)$ that maps the standard orthonormal basis e_1, e_2, e_3 to $v(t), n(t), b(t)$. Then $\eta(t) := g'(t)g(t)^{-1}$ is a tangent vector at the identity element of $SO(3)$. In other words, $\eta(t)$ belongs to the Lie algebra of $SO(3)$, which consists of skew-symmetric 3×3 real matrices. Therefore we can write

$$
g'(t) = \begin{pmatrix} 0 & \alpha(t) & \beta(t) \\ -\alpha(t) & 0 & -\tau(t) \\ -\beta(t) & \tau(t) & 0 \end{pmatrix} g(t)
$$

for functions $\alpha(t), \beta(t), \tau(t)$. But the definition of κ and n implies that $\alpha = \kappa$ and $\beta = 0$. □

The function $\tau(t)$ obtained in the previous proposition will be called the *torsion* of γ. Notice that it contains the information about the way the *osculating plane* (spanned by v and n) changes at each point. Writing

$$
B := \begin{pmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & -\tau \\ 0 & \tau & 0 \end{pmatrix},
$$

then γ specifies a curve $g(t) \in SO(3)$ that satisfies $g'(t) = B(t)g(t)$.

Conversely, an arbitrary choice of the functions κ and τ specifies a curve $\gamma(t)$ (uniquely up to an isometry of $\mathbb{R}^3$) whose curvature and torsion are κ and τ, respectively. This is the content of the next exercise.

Exercise 5.6.2. Let $g(t)$ be the unique solution of the differential equation $g'(t) = B(t)g(t)$ such that $g(0)$ is the identity element of $SO(3)$. Let $v(t)$ be the first column vector in $g(t)$ and define $\gamma(t) = p_0 + \int_0^t v(s)ds$. Show that $\gamma(t)$ is curve in $\mathbb{R}^3$ with curvature κ and torsion τ. Show that if $\xi(t)$ is another curve with the same curvature and torsion, then there exists a fixed $f \in E(3)$ such that $\xi = f \circ \gamma$.

The above discussion suggests the following control-theoretic point of view to describing curves in $\mathbb{R}^3$. Let X_1, X_2, X_3 be the right-invariant vector fields on $SO(3)$ such that

$$X_1(e) = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, X_2(e) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, X_3(e) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix},$$

where e is the identity element of $SO(3)$. Then the moving frame $g(t)$ along γ can be expressed as a solution of the differential equation on $SO(3)$ of the form

$$\frac{dg}{dt} = \kappa(t)X_1(g) - \tau(t)X_3(g).$$

Therefore we can view $t \mapsto u(t) = (\kappa(t), \tau(t)) \in \mathbb{R}^2$ as a control function that directs the curve to bend about itself in space according to preassigned functions of curvature and torsion.

If we let κ and τ be functions both of t and $\omega \in \Omega$, where $(\Omega, \mathcal{F}, P)$ is a probability space, in such a way that for each ω these are Lipschitz continuous functions of t, then for each ω the unique solution $X_t(\omega) := p(t, \omega, p_0)$ through a given point p_0 defines a stochastic process with values in M. With some work, it will be seen later how to obtain a process when the functions κ and τ are only square integrable. I invite you to imagine the shape of a curve in space for which κ and τ take only the values 1 and -1, which alternate after a random (exponential) time.

Later on it will be seen how to interpret the Serret-Frenet equations when the curvature and torsion functions are such things as the "derivative" of Brownian motion.

c. The Main Technical Results. We recall that $\mathcal{P}_p(M)$ denotes the space of all continuous curves $\gamma : [0, \infty) \to M$ such that $\gamma(0) = p$. Our goal is to solve the differential equation

$$\dot{p} = X_0(p) + \sum_{k=1}^{d} \dot{\omega}_i(t) X_k(p), \tag{5.2}$$

with initial condition $p(t) = p_0$, where $X_0, X_1, \dots, X_d$ are smooth vector fields on M and $\omega(t) = (\omega_1(t), \dots, \omega_d(t))$ is a curve in $\mathcal{P}_0(\mathbb{R}^d)$. We represent the solution by $p(t, p_0, \omega)$.

When ω is a differentiable curve, so that its time derivative $\dot{\omega}$ is continuous, solving Equation 5.2 only invoves standard techniques in ordinary differential equations, but clearly the meaning of "solving" 5.2 will have to be clarified in the general case. Before explaining this point, we describe an approximation scheme that will produce the solution in the smooth case and will also be used in the non-smooth case.

CHAPTER 6

Riemannian Manifolds

1. Riemannian Metrics

A *Riemannian metric* on a differentiable manifold M is a correspondence which associates to each point $p \in M$ an inner product $\langle \cdot, \cdot \rangle_p$ (that is, a bilinear, positive-definite form) on the tangent space T_pM, which varies differentiably in the following sense: For any pair of differentiable vector fields, X, Y, on M, $p \mapsto \langle X(p), Y(p) \rangle_p$ is a differentiable function.

Let M and N be Riemannian manifolds. A diffeomorphism $f : M \to N$ (that is, a differentiable bijection with a differentiable inverse) is called an *isometry* if $\langle u, v \rangle_p = \langle df_p u, df_p v \rangle_{f(p)}$, for all $p \in M$ and all $u, v \in T_pM$. More generally, if $f : M \to N$ is a differentiable immersion of a manifold M into a manifold N (this means that df_p is injective for each $p \in M$) and N has a Riemannian metric $\langle \cdot, \cdot \rangle$, then the metric can be *pulled-back* to a Riemannian metric $f^* \langle \cdot, \cdot \rangle$ on M under the definition

$$f^* \langle u, v \rangle_p := \langle df_p u, df_p v \rangle_{f(p)}.$$

If both M and N are Riemannian and the Riemannian metric on M coincides with the pulled-back metric on N uner an immersion $f : M \to N$, then we say that f is an *isometric immersion*.

The Riemannian metric on M allows one to define the length of a piecewise differentiable curve $\gamma : [t_1, t_2] \to M$ as

$$L[\gamma] := \int_{t_1}^{t_2} \|\gamma'(t)\| dt,$$

where $\|\gamma'(t)\| = \langle \gamma'(t), \gamma'(t) \rangle$. It is often convenient to consider the action functional of the curve (called the *energy of the curve* in Riemannian geometry books):

$$E[\gamma] = \int_{t_1}^{t_2} \|\gamma'(t)\|^2 dt.$$

The *distance* between two points p, q of a Riemannian manifold is defined as the infimum of the lengths of all piecewise differentiable curves $\gamma : [t_1, t_2] \to M$ such that $\gamma(t_1) = p$ and $\gamma(t_2) = q$.

Exercise 6.1.1. Show that the distance just defined actually turns M into a metric space.

1. *Conformally Euclidian Metrics and Optical Length.* Let $M = \mathbb{R}^n$ and denote by $\langle \cdot, \cdot \rangle_0$ the ordinary dot product, which defines the standard Riemannian structure in $\mathbb{R}^n$. Let $U \subset \mathbb{R}^n$ be an open subset. A metric $\langle \cdot, \cdot \rangle$ on U is said to be *conformally Euclidian* if it can be written as

$$\langle \cdot, \cdot \rangle := e^h \langle \cdot, \cdot \rangle_0$$

where $h : U \to \mathbb{R}$ is a differentiable function.

An interesting example of such a metric occurs in geometrical optics. The energy carried by a beam of light propagates in space with speed $v = c/n$, where c is the speed of light in vacuum (a constant) and n is a positive function called the *refractive index*, which is determined by the electric and magnetic characteristics of the medium in which light propagates. We write $n = e^h$. The *optical length* of a curve γ is the length of γ with respect to the conformally Euclidian metric $e^{2h}\langle\cdot,\cdot\rangle$. The optical length is then the product of c by the time light would take to follow the path γ. (The parametrization $t \mapsto \gamma(t)$ is arbitrary and t is not being regarded as time here.)

The *principle of Fermat* asserts that the optical length of the actual path taken by light between two points $p_1, p_2 \in U \subset \mathbb{R}^3$ is less than or equal to the optical length of any curve which joins p_1 and p_2 (at least as long as the points are sufficiently close to each other to belong to a neighborhood where length minimizing curves between any two points are unique. More on this condition later). In other words, actual light rays are geodesics for the optical length metric on U.

2. *Submanifolds of* $\mathbb{R}^n$. Let M be a differentiable submanifold of $\mathbb{R}^n$. Then the ordinary dot product, $\langle\cdot,\cdot\rangle_0$, immediately yields a Riemannian metric on M simply by its restriction to each T_pM, where the tangent space is regarded as a linear subspace of $\mathbb{R}^n$.

3. *Lie Groups.* Let $\langle\cdot,\cdot\rangle_e$ be an arbitrary positive inner product on $\mathfrak{g} = T_eG$. For each $g \in G$, denote by $L_g : G \to G$ the left-translation diffeomorphism, defined by $h \mapsto gh$. Define an inner-product $\langle\cdot,\cdot\rangle_g$ on T_gG by pulling-back the one at the identity element via the map $L_{g^{-1}}$ (which sends g to e). In other words, set

$$\langle u, v\rangle_g := \langle (dL_g)_e^{-1}u, (dL_g)_e^{-1}v\rangle_e.$$

This defines a Riemannian metric on G which has the property of being *left-invariant*, that is

$$L_g^*\langle\cdot,\cdot\rangle_h = \langle\cdot,\cdot\rangle_{g^{-1}h}$$

for all $g, h \in G$. Therefore left translations are isometries.

As may be expected, problems about the Riemannian geometry of Lie groups with left-invariant metrics reduce by-and-large to linear algebra questions on the Lie algebra of G. We will see instances of this idea later on. Of course, all that has been said applies equally well to right-invariant metrics. It is, however, not always possible to define a metric on G that is both right and left-invariant.

EXERCISE 6.1.2. Show that if a connected Lie group G admits a Riemannian metric that is both right and left-invariant, then G has the following property. Let $Z = \{g \in G : gh = hg \text{ for al } h \in G\}$ be the center of G. (This is clearly a closed abelian normal subgroup of G.) Then the quotient group G/Z is compact.

give details for this exercise

2. Connections

Let $X = h_1\frac{\partial}{\partial x_1} + \cdots + h_n\frac{\partial}{\partial x_n}$ be a differentiable vector field on $\mathbb{R}^n$ and $v \in T_p\mathbb{R}^n$ a tangent vector at p. Then the operation

$$D_vX := \lim_{t\to 0}\frac{X(p+tv) - X(p)}{t} = \sum_{i=1}^{n}(vh_i)(p)\frac{\partial}{\partial x_i} \tag{6.1}$$

may be regarded as a directional derivative of X in the direction v.

The definition of D_vX makes use of the affine structure of $\mathbb{R}^n$, as it requires translating a vector from one tangent space to another between taking the difference. To see how such an operation can be introduced on a general manifold it will help to look at some of the properties of D_v. The main general properties are stated in the next exercise.

EXERCISE 6.2.1. Show that the operation D defined in 6.1 has the following properties and that these properties (up to item 5) uniquely characterize D. Here X, Y, Z are vector fields, v, u are tangent vectors at a point p, $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable function, and $\langle \cdot, \cdot \rangle$ is the dot product.

(1) $D_v(X + Y) = D_vX + D_vY$
(2) $D_v(fX) = (vf)X + fD_vX$
(3) $D_{au+bv}X = aD_uX + bD_vX$
(4) $v\langle X, Y\rangle = \langle D_vX, Y\rangle + \langle X, D_vY\rangle$
(5) $D_XY - D_YX = [X, Y]$
(6) $[D_X, D_Y]Z = D_{[X,Y]}Z$.

4. *Hypersurfaces in* $\mathbb{R}^{n+1}$. Let M be a hypersurface in $\mathbb{R}^{n+1}$, that is, an n-dimensional smooth submanifold. Denote by $\Pi_p : \mathbb{R}^{n+1} \to T_pM$ the orthogonal projection map. If $N(p)$ is a vector at p perpendicular to the hypersurface, then $\Pi_p u = u - \langle u, N(p)\rangle N(p)$.

Give M the Riemannian metric given by restricting the ordinary dot-product of $\mathbb{R}^{n+1}$ to each T_pM. Let D be as above and define, for a vector field X on M and $v \in T_pM$, the operation

$$\nabla_v X := \Pi_p D_v X = D_vX - \langle D_vX, N(p)\rangle N(p).$$

EXERCISE 6.2.2. Show that the ∇ just defined satisfies properties 1 through 5 of the previous exercise.

5. *Connections on Manifolds.* Let now M be a differentiable manifold. A *connection* on (the tangent bundle of) M is an operation ∇ that associates to each differentiable vector field X and tangent vector $v \in T_pM$ another vector $\nabla_vX \in T_pM$ having the following properties:

$$\nabla_{au+bv}X = a\nabla_uX + b\nabla_vX \tag{6.2}$$

$$\nabla_v(X + Y) = \nabla_vX + \nabla_vY \tag{6.3}$$

$$\nabla_v(fX) = (vf)\nabla X + f\nabla_vX. \tag{6.4}$$

A connection ∇ is said to be *torsion-free* if, moreover,

$$\nabla_XY - \nabla_YX = [X, Y]. \tag{6.5}$$

Suppose now that M is provided with a Riemannian metric $\langle \cdot, \cdot \rangle$. We say that ∇ is a *metric connection* if for all $v \in T_pM$ and vector fields X, Y,

$$v\langle X, Y\rangle = \langle \nabla_vX, Y\rangle + \langle X, \nabla_vY\rangle. \tag{6.6}$$

It is interesting to remark that if ∇^1 and ∇^2 are two connections on a manifold M, then $(X, Y) \mapsto T(X, Y) = \nabla^1_XY - \nabla^2_XY$ is a tensor field of type $(2, 1)$. This means that it is a multilinear function of its (two) vector field arguments, the result of the operation is itself a vector field and it is a "purely algebraic operation" in that no derivatives are really involved in it. More precisely, given a function f on M and vector fields X, Y we have $T(fX, Y) = fT(X, Y) = T(X, fY)$.

Conversely, $\nabla'_X Y = \nabla_X Y + T(X,Y)$ is a connection on M whenever ∇ is a connection and T is a $(2,1)$ tensor field.

EXERCISE 6.2.3. Let ∇ and ∇' be two connections on a manifold M. Show:

(1) If both ∇ and ∇' are torsion-free, then $T = \nabla - \nabla'$ is a symmetric tensor, i.e., $T(u,v) = T(v,u)$ for all vectors $u, v \in T_pM$.
(2) If M is Riemannian and ∇ and ∇' are metric connections then, for all vectors $u, v, w \in T_pM$, $\langle T(u,v), w\rangle + \langle u, T(v,w)\rangle = 0$.
(3) If ∇ and ∇' are both torsion-free and metric for the same Riemannian metric, then $T = 0$. In other words, there can be at most one torsion-free metric connection on a Riemannian manifold.

PROPOSITION 6.2.4 (The Levi-Civita connection). Let $(M, \langle\cdot,\cdot\rangle)$ be a Riemannian manifold. Then there exists on M one and only one torsion-free metric connection.

PROOF. Uniqueness is shown in the previous exercise. To show existence, define $\nabla_X Y$ by the equation:

$$2\langle\nabla_X Y, Z\rangle = X\langle Y,Z\rangle - Z\langle X,Y\rangle + Y\langle X,Z\rangle + \langle Y,[X,Z]\rangle + \langle X,[Z,Y]\rangle - \langle Z,[Y,X]\rangle.$$

It is now a direct verification that ∇ actually defines a torsion-free metric connection. $\square$

Property (6) of Exercise 6.2.1 says that the ordinary Riemannian metric on $\mathbb{R}^n$ is *flat*. This concept will be discussed later. We only make here the remark contained in the next exercise.

EXERCISE 6.2.5. Suppose that $\langle\cdot,\cdot\rangle$ is a Riemannian metric on an open set $U \subset \mathbb{R}^n$ and let ∇ be its Levi-Civita connection. Also suppose that

$$\nabla_X\nabla_Y - \nabla_Y\nabla_X = \nabla_{[X,Y]}.$$

Then there exists a diffeomorphism $\Phi : U \to V \subset \mathbb{R}^n$ that pulls the dot product to the given Riemannian metric. In other words, U with the given Riemannian metric is isometric to an open subset of $\mathbb{R}^n$ with the ordinary metric.

elaborate on this exercise. Give more details of proof.

6. *Lie Groups.* Let G be a Lie group. We recall that the Lie algebra, $\mathfrak{g}$, of G is the linear space of all left-invariant vector fields on G. Since a left-invariant vector field is uniquely specified by its value at the unit element e, we can identify $\mathfrak{g}$ with the tangent space T_eG, although it will be more useful now not to make this identification.

EXERCISE 6.2.6 (The Torsion Tensor of a Connection). Let ∇ be a connection on a differentiable manifold M.

(1) Show that $T(X,Y) := \nabla_X Y - \nabla_Y X - [X,Y]$ is a $(2,1)$-tensor field, and that $T(X,Y) = -T(Y,X)$. It is called the *torsion tensor* of ∇.
(2) Show that $\nabla_X Y := \nabla_X Y - \frac{1}{2}T(X,Y)$ is a torsion-free connection.

EXERCISE 6.2.7. Show that the condition $\nabla u X = 0$ for any left-invariant vector field X on G specifies a unique connection on G. Check that this is a torsion-free connection if and only if the Lie algebra of G is abelian. More precisely, show that if $X, Y \in \mathfrak{g}$, then $T(X,Y) = -[X,Y]$.

EXERCISE 6.2.8. Let G be a Lie group and $\langle\cdot,\cdot\rangle$ a left-invariant Riemannian metric on G. Show that the Levi-Civita connection on G satisfies the following: for left-invariant vector fields X, Y, Z,

$$2\langle\nabla_X Y, Z\rangle = \langle[X,Y],Z\rangle - \langle[X,Z],Y\rangle - \langle X,[Y,Z]\rangle.$$

3. Covariant Derivatives

Let M be a differentiable manifold and ∇ any connection on M. Let $\alpha : (a,b) \to M$ be a differentiable curve. A differentiable vector field along α is a map $X : (a,b) \to TM$ such that $X(t) \in T_{\alpha(t)}M$ for each $t \in (a,b)$.

Define the covariant derivative of X, denoted $\frac{\nabla X}{dt}$, as follows. Let $X_1, \ldots, X_n$ be a local frame and write $X(t) = \sum_{i=1}^n h_i(t)X_i(\alpha(t))$. Then

$$\frac{\nabla X}{dt}(t) = \sum_{i=1}^n (h_i'(t)X_i(\alpha(t)) + h_i(t)\nabla_{\alpha'(t)}X_i).$$

EXERCISE 6.3.1. Show that $\frac{\nabla X}{dt}$ is well-defined, that is, it does not depend on the choice of the frame X_i. Furthermore, if $X(t)$ already is the restriction to α of a vector field on M, show that $\frac{\nabla X}{dt}(t) = \nabla_{\gamma'(t)}X$.

EXERCISE 6.3.2. Show that the covariant derivative satisfies the following properties:

(1) $\frac{\nabla}{dt}(aX + bY) = a\frac{\nabla X}{dt} + b\frac{\nabla Y}{dt}$
(2) $\frac{\nabla}{dt}(gX) = g'X + g\frac{\nabla X}{dt}$, for any differentiable function $g : (a,b) \to \mathbb{R}$.
(3) If ∇ is a metric (Riemannian) connection, then

$$\frac{d}{dt}\langle X,Y\rangle = \left\langle \frac{\nabla X}{dt}, Y\right\rangle + \left\langle X, \frac{\nabla Y}{dt}\right\rangle.$$

Given a a (twice) differentiable curve $\alpha : (a,b) \to M$, the covariant derivative, $\frac{\nabla\alpha'}{dt}$, of the velocity of α is called the *acceleration* of the curve. A curve that has zero acceleration is called a *geodesic* (relative to ∇). Notice that no Riemann metric is needed to define geodesics in this generality. Of course, it only makes sense to speak about the length minimizing property of geodesics for metric connections. We will return later to the variational characterization of geodesics on a Riemannian manifold.

discuss existence, uniqueness of geodesic equation; completeness

EXERCISE 6.3.3. Let M be a Riemannian manifold and $\alpha : (a,b) \to M$ a differentiable curve parametrized by arclength. Show that the acceleration of α is aways perpendicular to the velocity.

EXERCISE 6.3.4. Let e_1, e_2, e_3 denote the standard basis of $\mathbb{R}^3$ and S^3 the sphere of unit radius centered at the origin. The intersection of the sphere with the affine plane $\{(x,y,z) \in \mathbb{R}^3 : z = s\}$, for $-1 < s < 1$, is a circle which, parametrized by arclength, can be written as

$$\alpha(t) = \sqrt{1-s^2}\left(\cos\frac{t}{\sqrt{1-s^2}}e_1 + \sin\frac{t}{\sqrt{1-s^2}}e_2\right) + se_3.$$

Show that the acceleration of the curve (for the Levi-Civita connection) is

$$\frac{\nabla\alpha'(t)}{dt} = -\frac{s^2}{1-s^2}\left(\cos\frac{t}{\sqrt{1-s^2}}e_1 + \sin\frac{t}{\sqrt{1-s^2}}e_2\right) + se_3.$$

Notice, in particular, that the curve is a geodesic if $s = 0$ and that the acceleraton goes to infinity as s approaches ± 1.

7. *Geodesics on Lie Groups.* Suppose that G is a connected Lie group, $\langle \cdot, \cdot \rangle$ a left-invariant Riemannian metric, and let ∇ be the Levi-Civita connection on G.

EXERCISE 6.3.5. Show that ∇ is left-invariant. (Prove this by showing that the pulled-back connection under L_g, defined by $\nabla_X^{L_g} Y := (L_{g^{-1}})_* \nabla_{(L_g)_* X} (L_g)_* Y$, is also torsion-free and it is a metric connection for the pulled-back metric $(L_g)^* \langle \cdot, \cdot \rangle$. But the metric is left-invariant and the Levi-Civita connection is the unique torsion-free metric connection. Therefore $\nabla^{L_g} = \nabla$.)

EXERCISE 6.3.6. Show that the following properties are equivalent for a connected Lie group G with a left-invariant Riemannian metric $\langle \cdot, \cdot \rangle$.

(1) The metric bi-invariant.
(2) For all $X, Y, Z \in \mathfrak{g}$,
$$\langle [X, Z], Y \rangle + \langle X, [Y, Z] \rangle = 0.$$
(3) For all $X \in \mathfrak{g}$, $\nabla_X X = 0$.
(4) The curve $t \mapsto g(t)$ is a geodesic if and only if $g(t) = g\sigma(t)$, for some $g \in G$ and some one-parameter subgroup $\sigma(t)$. (A one-parameter subgroup is a homomorphism from the additive one-dimensional group $\mathbb{R}$ into G.)

explain details of exercise

It was claimed in an earlier exercise that a connected Lie group G has a bi-invariant metric if and only if G/Z is compact, where Z is the center. In particular, the previous exercise implies that the geodesics of the orthogonal group $O(n)$ are the curves of the form $t \mapsto Ae^{tB}$, where A is a matrix in $O(n)$ and B is a skew-symmetric $n \times n$ matrix.

a. Parallel Translation. A vector field X along a differentiable curve α is said to be *parallel* (relative to a connection ∇) if $\frac{\nabla X}{dt} = 0$.

For example, if $X(t) = \sum_{i=1}^n h_i(t) \frac{\partial}{\partial x_i}$ and D is the Euclidian connection in $\mathbb{R}^n$, then $\frac{DX}{dt} = \sum_{i=1}^n {h'}_i \frac{\partial}{\partial x_i} = 0$ implies that h_i is constant for each i, so X is a constant vector field.

PROPOSITION 6.3.7 (Parallel Translation). Let $\alpha : [a, b] \to M$ be a differentiable curve and $v \in T_{\alpha(a)} M$, then there exists a unique vector field X along α such that $\frac{\nabla X}{dt}(t) = 0$ and $X(a) = v$.

PROOF. It suffices to consider the case in which $\alpha([a, b])$ is contained in a coordinate neighborhood, so that we have a frame of vector fields $X_1, \ldots, X_n$. Write $\nabla_{X_i} X_j = \sum_{k=1}^n \Gamma_{ij}^k X_k$. Also write $X = \sum_{i=1}^n h_i X_i$, $\alpha'(t) = \sum_{i=1}^n a_i(t) X_i(\alpha(t))$ and $v = \sum_{i=1}^n v_i X_i(\alpha(a))$. Then we have $\frac{\nabla X}{dt} = \sum_{k=1}^n ({h'}_k + \sum_{i,j=1}^n h_i a_j \Gamma_{ij}^k) X_k$. Therefore the initial value problem $\frac{\nabla X}{dt} = 0$, $X(a) = v$ is equivalent to the system of equations, where $k = 1, \ldots, n$:

$$\begin{aligned} {h'}_k + \sum_{i,j=1}^n a_j(t) \Gamma_{ij}^k(\alpha(t)) h_i = 0 \\ h_k(0) = v_k. \end{aligned}$$

Let h' be the column vector with components ${h'}_i$ and define a matrix $\eta(t)$ with entries $\eta_{ki}(t) := -\sum_{j=1}^n a_j(t)\Gamma_{ji}^k(\alpha(t))$. Then the system takes the form

$$h' = \eta(t)h$$
$$h(0) = (v_1, \ldots, v_n)^t.$$

This is a system of linear differential equations with initial value. By the general theory of ODEs it must have a unique solution which is defined over the whole intervale $[a, b]$. $\square$

Notice that since the equation of parallel transport is linear the map $v \in T_{\alpha(a)}M \mapsto P_\alpha v \in T_{\alpha(b)}M$ that associates to each vector its parallel translation to the endpoint of α is a linear map. In particular, if α is a closed curve with initial and end point at p, then P_α is a linear transformation of the vector space T_pM. It is, in fact, an invertible transformation and it is easy to verify (by uniqueness of solutions of the differential equation) that P_γ^{-1} is parallel translation along the same curve traverse from end to beginning.

EXERCISE 6.3.8. Let $\mathcal{C}_p(M)$ denote the set of piecewise differentiable closed curves beginning and ending at p. Given γ_1, γ_2 in $\mathcal{C}_p(M)$ we define the concatenation $\gamma_2 \cdot \gamma_1$ to be the curve that traverses γ_1 followed by γ_2, and by γ^{-1} the curve that corresponds to traversing γ in reverse. Show the following:

(1) $P_{\gamma^{-1}} = P_\gamma^{-1}$;
(2) $P_{\gamma_2 \cdot \gamma_1} = P_{\gamma_2} \circ P_{\gamma_1}$;
(3) the set of all P_γ for $\gamma \in \mathcal{C}_p(M)$ constitutes a group, called the *Holonomy group* of the connection ∇;
(4) if ∇ is a Riemannian connection (possibly with torsion) for some Riemannian metric $\langle \cdot, \cdot \rangle$ on M, then P_γ lies in the group of orthogonal transformations, $O(T_pM)$, consisting of linear isometries of T_pM. If γ is a general continuous curve, then P_γ is a linear isometry from $T_{\gamma(a)}M$ to $T_{\gamma(b)}M$.

b. The Riemannian Manifold of a Mechanical System. We can describe the motion of a mechanical system in $\mathbb{R}^3$ as follows. Let B, the *body*, be a topological space, possibly with some further structure given to it. For example, B could be of finite graph, a metric space, a Riemannian manifold, and so forth. For example, suppose that B is the union of (possibly intersecting) subsets $B_i \subset \mathbb{R}^3$ such that each B_i has the metric induced from $\mathbb{R}^3$. (This might correspond to a linked system of solid bodies.) Then a configuration of B would be a bijection from B to a subset of $\mathbb{R}^3$ whose restriction to each B_i is an isometry onto its image. The set B may also be given functions (scalar functions, tensor fields, etc.) that describe intrinsic properties of the body such as mass or charge distribution, elastic properties, etc.

A *configuration* of the body is a map $\alpha : B \to \mathbb{R}^3$ which is a bijection between B and $\alpha(B)$ and $\alpha : B \to \alpha(B)$ is an isomorphism for the given structure on B and the corresponding structure on $\alpha(B)$ that results by restricting to this set the appropriate structure of $\mathbb{R}^3$. For example, if B is a subset of $\mathbb{R}^3$ that we wish to regard as a rigid body, then the structure on B is the metric space structure induced from the Euclidian metric in $\mathbb{R}^3$ and a configuration α is an isometry between B and $\alpha(B)$.

Let M denote the set of all configurations of B. This is naturally a topological space with the compact open topology, and the topology is induced by the following family of metrics: if K is any compact subset of B, set

$$d_K(\alpha, \beta) = \sup_{p \in K} |\alpha(p) - \beta(p)|.$$

It will be convenient to introduce the *position map*

$$\Phi : M \times B \to \mathbb{R}^3$$

which associates to each $p \in B$ and $\alpha \in M$ the image $\alpha(p)$.

A *motion* of the body B is a continuous path in the configuration space, that is, a continuous map $\gamma : I \to M$, parametrized by some interval $I \subset \mathbb{R}$.

We suppose from now on that M has a is a differentiable manifold structure for which the position map is smooth in the first variable, that is, for each $p \in B$, $\alpha \mapsto \Phi(\alpha, p)$ is a differentiable map. If this is the case, then each tangent vector $\xi \in T_\alpha M$ specifies a vector field on $\alpha(B)$. Indeed, the value of the vector field at a point $\alpha(p)$ is the velocity of the curve $t \mapsto \Phi(\alpha(t), p)$ where $\alpha(t)$ is a path in M such that $\alpha(0) = \alpha$ and $\alpha'(0) = \xi$. The value of the vector field at p will be written ξ_p.

We suppose that M is a differentiable manifold and that B is a fixed compact subset of $\mathbb{R}^3$ which is given a density of mass function ρ (possibly a distribution). We give M the following Riemannian metric: for $\xi, \eta \in T_\alpha M$

$$\langle \xi, \eta \rangle := \frac{1}{m} \int_B \langle \xi_p, \eta_p \rangle \rho(p) d\lambda(p),$$

where m is the total mass of B (the integral of ρ over B), λ indicates the Lebesgue measure in $\mathbb{R}^3$, and $\langle \xi_p, \eta_p \rangle$ is the ordinary dot product.

We say that elements of the tangent bundle of M are *states* of the mechanical system. Given a state $\xi \in T_\alpha M$, the total kinetic energy of the body in (that state) is given by

$$\frac{1}{2} \|\xi\|^2 = \frac{1}{2m} \int_B |\xi_p|^2 \rho(p) d\lambda(p).$$

In the general scheme of Lagrangian mechanics, the actual motion of a conservative mechanical system with potential $U : M \to \mathbb{R}$ corresponds to a critical point $\alpha \in M$ of the action functional

$$A[\gamma] := \int_{t_0}^{t} \left(\frac{1}{2} \|\gamma'(s)\|^2 - U(\gamma(s)) \right) ds$$

with, say, $\gamma(t_0)$ and $\gamma'(t_0)$ specified. In the absence of forces and potentials (besides those implicit in the geometric constraints imposed on B), $U = 0$ and a critical path of A is nothing but a geodesic for the Riemannian metric on M. We discuss later how to incorporate forces in this picture.

this should go in the next chapter after general discussion about Riemannian manifolds

c. Controling the Motion of a Rigid Body. In Newtonian mechanics, the motion of a particle of mass m under a force f is determined by Newton's equations. We first describe these equations under the assumption that the x_i define an *inertial coordinate system*, that is, a coordinate system such that any particle that has zero acceleration with respect to it is to be regarded as not being acted upon by any (nonzero) forces. (For us, earthlings, a frame fixed at the center of the sun that does not rotate with respect to the "fixed stars" is considered by the physicists to be a good enough such a system.)

Now, Newton's equations say that if a particle at time t has position coordinates $x = (x_1, x_2, x_3)$, velocity coordinates $v = (v_1, v_2, v_3)$, and is subject at that point to a force $f = f_1(x,v,t)e_1 + f_2(x,v,t)e_2 + f_3(x,v,t)e_3$, then the rate of change of linear momentum $q = mv$ at time t is $\dot{q}_i = f_i(x,v,t)$. If m does not depend on t, then $m\dot{v} = f(x,v,t)$.

We would like now to express these equations in non-inertial frames. Let e_1, e_2, e_3 be the standard orthonormal frame of $\mathbb{R}^3$ and $E_1(t), E_2(t), E_3(t)$ another orthonormal frame that is moving in space, having base point $p_0(t) \in \mathbb{R}^3$. (We think of $E_i(t)$ as a vector in $T_{p_0}\mathbb{R}^3$.) Let x_i be the standard coordinates in $\mathbb{R}^3$ and denote by X_i the new coordinates for which $X_i(p_0) = 0$ and $dX_i(E_j) = \delta_{ij}$. The two coordinate systems are related by $x_i = \sum_j O_{ij}(t)X_j + c_i(t)$, where $O(t) := (O_{ij}(t))$ is an orthogonal matrix. Write X, V, Q, F, etc, for the (column) vectors in the moving coordinates of the position, velocity, momentum, force, etc. Notice that F is a vector at X and $f(x,v,t) = O(t)F(X,V,t)$. Then $x = O(t)X + c(t)$. Differentiating, and writing $\Omega = O^{-1}O'$ and $\zeta = O^{-1}c'$, $v = O'(t)X + OV + c'(t) = O(t)(V + \Omega(t)X + \zeta)$, we get

$$q = O(t)(Q + \Omega(t)mX + m\,\zeta).$$

It follows that

$$O(t)F(X,V,t) = O'(t)(Q+\Omega(t)mX+m\,\zeta)+O(t)(\dot{Q}+\Omega'(t)mX+\Omega(t)Q+\frac{d}{dt}(m\,\zeta)).$$

Write $\frac{D}{dt} = \frac{d}{dt} + \Omega$. Then

$$F = \frac{D}{dt}(m\,\zeta + \frac{D}{dt}(mX)).$$

Suppose that we would like now to studey the motion of a rigid body. A point of the body is held fixed at a given position, which we assume to be the origin and the body is allowed to rotate about the origin under the effect of a system of forces.

We choose a moving frame $E_1(t), E_2(t), E_3(t)$ that is kept fixed to the body and rotates along with it. Therefore the coordinate vector X of each point of the body is constant and $V = 0$. We also have in this case that $c(t)$ is constant, so $\zeta = 0$. Let φ denote the density of force and ρ the density of mass of the body, both expressed in the moving frame coordinates. (The density of mass will be supposed independent of time). We allow φ to be a distribution, say, a finite number of Dirac delta functions on the boundary of the body (plus the internal forces that keep the shape of the body unchanged). Therefore, as $\frac{D}{dt}X = \Omega X$, we have the following form for Newton's equation (expressed in terms of the densities):

$$\varphi = \rho\left(\frac{D}{dt}\right)^2 X = \rho\left(\frac{d}{dt} + \Omega\right)\Omega X = \left(\dot{\Omega} + \Omega^2\right)\rho X.$$

Our next goal is to express Newton's equation, which is now in the form $\varphi = \left(\dot{\Omega} + \Omega^2\right)\rho X$, as a differential equation on the (Lie algebra of the) orthogonal group.

The orthogonal group $SO(3)$ will be denoted G and its Lie algebra $\mathfrak{g}$. We can view $\mathfrak{g}$ as the tangent space of G at the identity element e, but also as the linear space of left-invariant vector fields on G. (A vector field Z on G is said to be left-invariant if for all $g \in G$, $(L_g)_*Z = Z$, where L_g is the diffeomorphism of G that maps $h \in G$ to gh.)

It will be convenient to define a norm on $\mathfrak{g}$ in the following way. Let $\xi \in \mathfrak{g}$, and write $\xi = g'(0)$, for a differentiable curve $g(t) \in G$ such that $g(0) = e$. The curve $g(t)$ corresponds to a motion of the rigid body. For each point of the body with coordinate vector (in the fixed frame) X, its motion in space is given by $x(t) = g(t)X$, and the velocity of that point at $t = 0$ is $v = \xi X$. (This is matrix multiplication of ξ by the column vector X.) Therefore, the kinetic energy of the body B, having density ρ and total mass m would be

$$\frac{1}{2}m\|\xi\|^2 := \int_B \frac{1}{2}\rho(X)|\xi X|^2 dX.$$

The norm $\|\xi\|$ comes from the inner product $\langle \xi, \eta \rangle$ which is given as follows (the inner product inside the integral sign is the ordinary dot product):

$$\langle \xi, \eta \rangle_e = \frac{1}{m}\int_B \rho(X)\langle \xi X, \eta X\rangle dX. \tag{6.1}$$

In general, if $\xi, \eta \in T_gG$, where g is not necessarily the identity element, we set

$$\langle \xi, \eta \rangle_g := \langle (dL_{g^{-1}})_g \xi, (dL_{g^{-1}})_g \eta \rangle_e.$$

This definition implies that if ξ and η are left-invariant vector fields on G, then $g \to \langle \xi(g), \eta(g)\rangle_g$ is constant in g.

In other words, we have used the kinetic energy of the moving body to introduce a (left-invariant) Riemannian metric on the Lie group G.

Introduce a (time dependent, left-invariant) vector field F on G by the following formula

$$\langle F, \xi \rangle := \int_B \langle \varphi, \xi X\rangle dX.$$

Since the right-hand side of the equation is a linear functional in ξ, the equation characterizes F by the duality between 1-forms and vector fields provided by the Riemannian metric. Note that the integral reduces to a sum of the form

$$\langle F, \xi \rangle = \sum_{k=1}^{l} F_k \cdot \xi X_k$$

when φ describes a system of Dirac forces applied to the points $X_1, \ldots, X_l$.

Exercise 6.3.9. Let ∇ denote the Levi-Civita connection on G. Show that ∇ is left-invariant. This requires the following definitions. Given a diffeomorphism $\Phi : M \to M$ of a manifold M (for example, $\Phi = L_g$, left-translation on a Lie group $M = G$) and a connection ∇ on M, the pull-back of ∇ under Φ is the connection ∇^Φ such that for vector fields X, Y gives

$$\nabla^\Phi_X Y = \Phi_*^{-1}\nabla_{\Phi_* X}\Phi_* Y.$$

Then ∇ on G is left-invariant if $\nabla^{L_g} = \nabla$ for all $g \in G$.

Exercise 6.3.10. Let G be a Lie group of dimension n with left-invariant Riemannian metric $\langle \cdot, \cdot \rangle$. Let $Z_1, \ldots, Z_n$ be an orthonormal basis of left-invariant vector fields. Denote by c_{ij}^k the structure constants of the Lie algebra, so that, by definition

$$[Z_i, Z_j] = \sum_{k=1}^{n} c_{ij}^k Z_k.$$

Denote by Γ_{ij}^k the Christoffel symbols of the Levi-Civita connection ∇, so, by definition,

$$\nabla_{Z_i} Z_j = \sum_{k=1}^{n} \Gamma_{ij}^k Z_k.$$

Show that

$$\Gamma_{ij}^k = \frac{1}{2}\left(c_{ij}^k + c_{ki}^j - c_{kj}^i\right).$$

EXERCISE 6.3.11. We know from a previous exercise that the correspondence

$$x = (x_1, x_2, x_3) \in \mathbb{R}^3 \mapsto \Omega(x) = \begin{pmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{pmatrix} \in \mathfrak{so}(3)$$

is a Lie algebra isomorphism, where $\mathbb{R}^3$ is given the Lie algebra bracket:

$$[x, y] = x \times y$$

(the ordinary cross product). If $\langle x, y\rangle_0$ represents the dot product of x and y and $\langle\cdot,\cdot\rangle$ is the inner product of $\mathfrak{so}(3)$ defined in Equation 6.1, show that there exists a positive symmetric 3×3 matrix L such that

$$\langle\Omega(x), \Omega(y)\rangle = \langle Lx, y\rangle_0.$$

The matrix L is called the *inertia operator*. Let u_1, u_2, u_3 be an orthonormal basis of $\mathbb{R}^3$ consisting of eigenvectors of the inertia operator, and call the eigenvalues I_1, I_2, I_3. Show that the Christoffel symbols of ∇ in the basis $\Omega(u_1), \Omega(u_2), \Omega(u_3)$ of $\mathfrak{so}(3)$ have the following form

$$\Gamma_{ij}^k = \frac{1}{2}\varepsilon_{ijk}\left(1 - \frac{I_i - I_j}{I_k}\right).$$

The symbol ε_{ijk} is equal to 1 (resp., -1) if i, j, k is an even (resp., odd) permutation of $1, 2, 3$.

EXERCISE 6.3.12. Let E_1, E_2, E_3 be a basis of left-invariant vector fields of $SO(3)$ such that $E_i = \Omega(u_i)$, where u_1, u_2, u_3 is an orthonormal basis of eigenvectors of the inertia operator with eigenvalues I_1, I_2, I_3. If $\gamma(t)$ is a geodesic of $SO(3)$ for the left-invariant metric of Equation 6.1 and $\gamma'(t) = \xi_1(t)E_1 + \xi_2(t)E_2 + \xi_3(t)E_3$, show that

$$\begin{aligned} I_1\dot{\xi}_1 &= (I_2 - I_3)\xi_2\xi_3 \\ I_2\dot{\xi}_2 &= (I_3 - I_1)\xi_3\xi_1 \\ I_3\dot{\xi}_3 &= (I_1 - I_2)\xi_1\xi_2. \end{aligned}$$

These are the so called *Euler equations* of motion of a rigid body.

EXERCISE 6.3.13. Let $G = SO(3)$ and give G the (left-invariant) Riemannian metric obtained from the kinetic energy of a rigid body, as in Equation 6.1. Let Z_1, Z_2, Z_3 be an orthonormal basis of left-invariant vector fields. Use Exercise 6.3.10 to show that

$$\Gamma_{ij}^k = \langle\nabla_{Z_i} Z_j, Z_k\rangle = \frac{1}{m}\int_B \rho(X)\langle Z_i Z_j X, Z_k X\rangle dX.$$

The notation $Z_i Z_j X$ should be read as the matrix product of (the constant matrices) Z_i, Z_j and the column vector X. (On the left-hand side of the equation Z_i is regarded as a left-invariant vector field on G while on the right-hand side Z_i is viewed as the skew-symmetric 3×3 matrix that corresponds to that left-invariant vector field.)

Let ∇ be the Levi-Civita connection on G and let $g(t)$ be a differentiable curve in G. We call $\frac{\nabla g'}{dt}$ the acceleration of g (with respect to the given connection). We claim that Newton's equation takes the form

$$m\nabla_t g' = F.$$

(The symbol $\nabla_t W := \nabla W/dt$ indicates the covariant derivative of a vector field $W(t)$ along the curve $g(t)$.)

To show this identity, let Z_1, Z_2, Z_3 be an orthonormal frame of left-invariant vector fields on G. Then

$$\begin{aligned}
\langle \nabla_t g'(t), Z_j\rangle &= \frac{d}{dt}\langle g'(t), Z_j\rangle - \langle g'(t), \nabla_t Z_j\rangle \\
&= \frac{d}{dt}\langle g'(t), Z_j\rangle - \left\langle g'(t), \sum_{i=1}^{3}\langle g'(t), Z_i\rangle \nabla_{Z_i} Z_j \right\rangle \\
&= \frac{d}{dt}\langle g'(t), Z_j\rangle - \sum_{i=1}^{3}\langle g'(t), Z_i\rangle\langle g'(t), \nabla_{Z_i} Z_j\rangle \\
&= \frac{d}{dt}\langle g'(t), Z_j\rangle - \sum_{i,k=1}^{3}\langle g'(t), Z_i\rangle\langle g'(t), Z_k\rangle \Gamma_{ij}^k \\
&= \frac{d}{dt}\langle \Omega(t), Z_j\rangle - \sum_{i,k=1}^{3}\langle \Omega(t), Z_i\rangle\langle \Omega(t), Z_k\rangle \Gamma_{ij}^k \\
&= \langle \dot{\Omega}(t), Z_j\rangle - \sum_{i,k=1}^{3}\langle \Omega(t), Z_i\rangle\langle \Omega(t), Z_k\rangle \frac{1}{m}\int_B \rho(X)\langle Z_i Z_j X, Z_k X\rangle dX \\
&= \langle \dot{\Omega}(t), Z_j\rangle - \int_B \frac{\rho(X)}{m}\left\langle \sum_{i=1}^{3}\langle \Omega(t), Z_i\rangle Z_i Z_j X, \sum_{k=1}^{3}\langle \Omega(t), Z_k\rangle Z_k X \right\rangle dX \\
&= \langle \dot{\Omega}(t), Z_j\rangle - \int_B \frac{\rho(X)}{m}\left\langle \Omega(t) Z_j X, \Omega(t) X\right\rangle dX \\
&= \int_B \frac{\rho(X)}{m}\langle (\dot{\Omega}(t) + \Omega(t)^2)X, Z_j X\rangle dX \\
&= \frac{1}{m}\int_B \langle \varphi(t, X), Z_j X\rangle dX \\
&= \frac{1}{m}\langle F, Z_j\rangle.
\end{aligned}$$

Therefore, $\nabla_t g'(t) = \frac{1}{m}F$ as claimed.

Notice, for example, that if the forces are due to jets attached to the surface of the body firing a constant jet flow, then F is a time-independent left-invariant vector field.

4. The Curvature Tensor

Given any connection ∇ on a manifold M, define

$$R(X, Y)Z = [\nabla_X, \nabla_Y]Z - \nabla_{[X,Y]}Z$$

for vector fields X, Y, Z on M. R is a $(3, 1)$-tensor field, called the *curvature tensor* of ∇. In particular, the value of $R(X, Y)Z$ at a point p only depends on the values of the vector fields at p. This follows from the property $R(fX,Y)Z = R(X,fY)Z = R(X,Y)fZ = fR(X,Y)Z$, which is easy to check.

A connection is called *flat* if its curvature tensor vanishes. For example, if D is the Euclidian connection on $\mathbb{R}^n$, then it is flat, by Property 6 of Exercise 6.2.1.

EXERCISE 6.4.1 (Algebraic Symmetries of R). Let $(M, \langle\cdot,\cdot\rangle)$ be a Riemannian manifold and ∇ its Levi-Civita connection. Show the following properites:

(1) $\langle R(X,Y)Z,T\rangle = -\langle R(Y,X)Z,T\rangle$;
(2) $\langle R(X,Y)Z,T\rangle = -\langle R(X,Y)T,Z\rangle$.

for vector fields X, Y, Z, T.

An immediate consequence of the previous exercise is that if M has dimension two and $\{e_1, e_2\}$ is an orthonormal basis for T_pM, then the function

$$k(p) := \langle R(e_1,e_2)e_2, e_1\rangle_p$$

completely specifies the tensor R. This quantity, moreover, does not depend on the choice of local orthonormal frame, so it gives a globally defined function on M. The function K is called the *Gauss curvature* of the surface.

EXERCISE 6.4.2. Let M be equipped with a connection ∇. Suppose that each point of M has a neighborhood U in which the holonomy group (for curves contained in U) is trivial. Show that the curvature tensor R vanishes identically.

The previous exercise has a converse.

PROPOSITION 6.4.3. If the curvature tensor vanishes identically over the manifold M, each point p has a neighborhood where parallel translation only depends on the endpoints of the curve (in that neighborhood) and not on the curve itself. In particular, for each basis $\{v_1, \dots, v_n\}$ of T_pM, one obtains a local frame of vector fields $\{X_1, \dots, X_n\}$ such that $X_i(p) = v_i$ for each i.

add a proof

EXERCISE 6.4.4. Suppose that ∇ is the Levi-Civita connection for some Riemannian metric $\langle\cdot,\cdot\rangle$ on M and that the curvature tensor is identically zero. Show that there exists a local orthonormal frame of commuting vector fields. Show that it follows from this fact that M is locally isometric to Euclidian space.

a. Sectional Curvature. More generally, if $\{e_1, e_2\}$ is an orthonormal basis for a two dimensional subspace $V \subset T_pM$, we define the *sectional curvature* of $(M, \langle\cdot,\cdot\rangle)$ along the plane V at p as $K(V) := \langle R(e_1,e_2)e_2, e_1\rangle_p$. It is easy to check that this quantity does not depend on the choice of basis of V.

We give next some examples where K is calculated.

1. *Conformally Euclidian Surfaces.* Let M be an open subset of $\mathbb{R}^2$ and $f(x,y)$ a positive function on U. Define on U a Riemannian metric

$$\langle\cdot,\cdot\rangle = \frac{1}{f(x,y)^2}\langle\cdot,\cdot\rangle^{\text{euc}}.$$

We will calculate the Gauss curvature using the orthonormal frame of vector fields

$$e_1 = f(x,y)\frac{\partial}{\partial x}, e_2 = f(x,y)\frac{\partial}{\partial y}.$$

The result and some intermediate steps of the calculation are given in the next exercise.

EXERCISE 6.4.5. Show the following claims:

(1) $[e_1, e_2] = -f_y e_1 + f_x e_2$;
(2) $2\langle \nabla_{e_i} e_j, e_k \rangle = -\langle e_j, [e_i, e_k] \rangle + \langle e_i, [e_k, e_j] \rangle - \langle e_k, [e_j, e_i] \rangle$;
(3) The derivatives $\nabla_{e_i} e_j$ are given by:

$$\nabla_{e_1} e_1 = f_y e_2,\ \nabla_{e_1} e_2 = -f_y e_1,\ \nabla_{e_2} e_1 = -f_x e_2,\ \nabla_{e_2} e_2 = f_x e_1.$$

(4) $\langle R(e_1, e_2)e_2, e_1 \rangle = f(f_{xx} + f_{yy}) - (f_x^2 + f_y^2)$.
(5) Let $f(x, y) = a + bx + cy$, and U the the half-space consisting of points where $a+bx+cy > 0$. Show that the Gauss curvature is constant negative, given by

$$k(x, y) = -c^2 \left[1 + \frac{b}{c} + (\frac{b}{c})^2 \right].$$

In particular, the metric

$$\langle \cdot, \cdot \rangle^{\text{hyp}} := \frac{1}{y^2} \langle \cdot, \cdot \rangle^{\text{euc}}$$

has constant negative curvature -1.

2. *Surfaces of Revolution.* Let M be a surface in $\mathbb{R}^3$ which is the image of the map

$$\Phi : (\theta, t) \in \mathbb{R} \times (a, b) \mapsto (R(t) \cos \theta, R(t) \sin \theta, t),$$

for some function $R(t)$. In the next exercise we calculate the Gauss curvature of M regarded as a Riemannian submanifold of $\mathbb{R}^3$ with the Euclidian metric.

EXERCISE 6.4.6. Set $Y_1 = \Phi_* \frac{\partial}{\partial \theta}$ and $Y_2 = \Phi_* \frac{\partial}{\partial t}$.

(1) Show that

$$Y_1 = R'(t) \cos \theta \frac{\partial}{\partial x} + R'(t) \sin \theta \frac{\partial}{\partial y} + \frac{\partial}{\partial z}$$
$$Y_2 = -R'(t) \sin \theta \frac{\partial}{\partial x} + R'(t) \cos \theta \frac{\partial}{\partial y}.$$

(2) Show that

$$e_1 = \frac{1}{R(t)} Y_1$$
$$e_2 = \frac{1}{\sqrt{1 + (R'(t))^2}} X_2$$

is an orthonormal frame on M.

(3) Show that

$$[e_1, e_2] = \frac{1}{\sqrt{1 + (R'(t))^2}} \frac{R'(t)}{R(t)} e_1.$$

(4) Show that

$$\nabla_{e_1} e_1 = -\langle e_1, [e_1, e_2] \rangle e_2 = -\frac{R'(t)}{R(t)\sqrt{1 + (R'(t))^2}} e_2$$
$$\nabla_{e_2} e_1 = -\langle e_2, [e_1, e_2] \rangle e_2 = 0$$
$$\nabla_{e_1} e_2 = \langle e_1, [e_1, e_2] \rangle e_1 = \frac{R'(t)}{R(t)\sqrt{1 + (R'(t))^2}} e_1$$
$$\nabla_{e_2} e_2 = -\langle e_2, [e_1, e_2] \rangle e_1 = 0.$$

(5) Show that the Gauss curvature is

$$K(\theta,t) = -\frac{R''(t)}{R(t)(1+(R'(t))^2)}.$$

(6) A sphere of radius a can be parametrized (as a surface of revolution) using $R(t)=\sqrt{a^2-t^2}$. Show that the Gauss curvature is constant equal to $1/a^2$.

(7) A hyperboloid, viewed as a surface of revolution, can be parametrized using $R(t)=\sqrt{a^2+t^2}$. Show that its Gauss curvature at the points $(a\cos\theta, a\sin\theta, 0)$ is $K=-1/a^2$.

3. *Curvature of Lie Groups.*

Exercise 6.4.7. Let G be a Lie group with a left-invariant metric $\langle\cdot,\cdot\rangle$, and ∇ the Levi-Civita connection. Let $\{Z_1,\dots,Z_n\}$ be an orthonormal basis of left-invariant vector fields on G. Show that the sectional curvature of the plane spanned by Z_i, Z_j is

$$\langle R(Z_i,Z_j)Z_j,Z_i\rangle = -\frac{3}{2}\sum_{k=1}^{n}\left\{\frac{1}{2}\left[(c_{ij}^k)^2+(c_{kj}^i)^2+(c_{ik}^j)^2\right]+c_{ij}^k c_{jk}^i - c_{ij}^k c_{ik}^j + c_{jk}^i c_{ki}^j\right\}.$$

Using this equation, or a direct calculation, show that the curvature tensor of the connected two dimensional Lie group whose Lie algebra has an orthonormal basis $\{X,Y\}$ such that $[X,Y]=Y$ vanishes identically. (This is the Lie algebra of the unique, up to isomorphism, connected non-abelian two-dimensional Lie group.)

verify that the signs in previous exercise are correct!

Before calculating the curvature of other Lie groups, we digress about Killing vector fields.

4. *Killing Vector Fields.* A vector field X on the Riemannian manifold M is said to be a Killing vector field if the flow Φ_t of X is a one-parameter group of isometries. Therefore, we have the following characterization of Killing fields.

Proposition 6.4.8. Let X be a vector field on a Riemannian manifold M, with Riemannian metric $\langle\cdot,\cdot\rangle$. Denote by Φ_t the (local) flow of X. Then the following are equivalent:

(1) X is a Killing field;
(2) $\langle(\Phi_t)_*Y,(\Phi_t)_*Z\rangle = \langle Y,Z\rangle$ for all vector fields Y,Z;
(3) $X\langle Y,Z\rangle = \langle[X,Y],Z\rangle + \langle Y,[X,Z]\rangle$, for all vector fields Y,Z;
(4) $\langle\nabla_Y X,Z\rangle + \langle\nabla_Z X,Y\rangle = 0$, for all vector fields Y,Z.

Proof. Items 1 and 2 are equivalent by definition. Item 3 is simply the infinitesimal equivalent of 2, and is obtained from 2 by differentiation in t. (Recall that $[X,Y] = -\frac{d}{dt}|_{t=0}(\Phi_t)_*Y$. By integration one obtains 2 from 3.) The equivalence between 3 and 4 follows by using the expression of the Levi-Civita connection of Proposition 6.2.4. □

Exercise 6.4.9. Let G be a Lie group with a bi-invariant Riemannian metric $\langle\cdot,\cdot\rangle$.Let X,Y,Z be left-invariant vector fields, then

(1) Show that $\nabla_X Y = \frac{1}{2}[X,Y]$;
(2) Show that $R(X,Y)Z = \frac{1}{4}[[X,Y],Z]$;

(3) Show that the sectional curvature of the plane V generated by orthonormal left-invariant vector fields X, Y is $K(V) = \frac{1}{4}\|[X,Y]\|^2$. In particular, the sectional curvature of a Lie group with bi-invariant metric is non-negative and is zero if and only if V is an abelian two-dimensional Lie algebra.

EXERCISE 6.4.10. Find the sectional curvatures of the group $SO(n)$ with the bi-invariant metric given on the Lie algebra by

$$\langle A, B\rangle = -\mathrm{Trace}(A^t B) = \mathrm{Trace}(AB).$$

An orthonormal basis is given by the skew-symmetric matrices E_{ij} whose (i,j)-entry is $1/\sqrt{2}$ and all other entries (except the (j,i) entry) are zero.

5. *Spheres and Symmetric Spaces.* In all examples up till now, curvature was calculated by using an orthonormal frame. It is always possible to find local such frames, but unless they are in some way "natural" (as is the case for Lie groups, for example) it may be more convenient to proceed in a different way. We will illustrate a different approach by calculating the sectional curvatures of spheres.

The key ideas used here will apply to general symmetric spaces, as we will see later on. Another symmetric space we will take as example later on is the complex projective space. A symmetric space is a special type of a *Riemannian homogeneous space*. The latter is a Riemannian manifold that admits a transitive group of isometries. The existence of a transitive group of isometries will allow us to reduce the calculation of geometric quantities such as curvatures to linear algebra computations on the tangent space of a fixed, arbitrary, point. For example, suppose that we want to calculate the curvature tensor of M. If $f : M \to M$ is any isometry of M, then $df_p R_p(u,v)w = R_{f(p)}(df_p u, df_p v) df_p w$ holds. (See the next exercise.)

EXERCISE 6.4.11. Let $f : M \to M$ be an isometry of the Riemannian manifold M. Let ∇ be the Levi-Civita connection. Show that for any vector fields X, Y,

$$f_*(\nabla_X Y) = \nabla_{f_* X} f_* Y.$$

(Note: This can be done as follows. Define ∇^f by setting $\nabla^f_X Y = f_*^{-1}\nabla_{f_*X} f_* Y$. Show that ∇^f is also a metric connection with zero torsion, so $\nabla^f = \nabla$, by uniqueness of the Levi-Civita connection.) Use this fact to show that $df_p R_p(u,v)w = R_{f(p)}(df_p u, df_p v) df_p w$ holds. In particular, if V is a two dimensional subspace of T_pM, then $K(V) = K(df_p V)$.

Let $M = S^n_r$ be the sphere of radius r in $\mathbb{R}^{n+1}$ with center at the origin. We give M the induced metric as a submanifold of $\mathbb{R}^{n+1}$ The group $G = SO(n+1)$ acts transitively on M by isometries. We fix the point $p_0 = (0,\dots,0,r)$, at which our calculations will be made.

Let X be an element of the Lie algebra $\mathfrak{g}$ of G, hence X can be regarded, both as a left-invariant vector field on G and as a skew-symmetric $n+1$-by-$n+1$ matrix. Then X can be regarded as a vector field on M, as follows. Denote by $\varphi_t = e^{tX} \in G$ the one-parameter group of G generated by X. Then for each t, φ_t is an isometry of M, and we write

$$\tilde{X}_p := \frac{d}{dt}|_{t=0}\varphi_{-t}(p) = -Xp.$$

The notation Xp indicates the matrix product of X and the column matrix representing p in the ordinary $\mathbb{R}^{n+1}$ coordinates.

EXERCISE 6.4.12. For each skew-symmetric matrix $A \in \mathfrak{so}(n+1)$ and $u \in T_pM$, where p is an arbitrary point of M, show that $\nabla_u \tilde{A} = -Au + r^{-2}\langle Au, p\rangle p$.

It will be convenient to distinguish between two types of elements of $\mathfrak{g}$. Notice that each element of this Lie algebra can be written as follows:

$$\begin{pmatrix} L & -u \\ u^t & 0 \end{pmatrix} = \begin{pmatrix} L & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & -u \\ u^t & 0 \end{pmatrix} \in \mathfrak{k} + \mathfrak{m}. \tag{6.1}$$

Here we denote $\mathfrak{k}$ the Lie algebra of $\mathfrak{so}(n)$, regarded as a Lie subalgebra of $\mathfrak{g}$, and $\mathfrak{m}$ is the n-dimensional subspace of $\mathfrak{g}$ consisting of matrices of the form $\begin{pmatrix} 0 & -u \\ u^t & 0 \end{pmatrix}$, where $u \in \mathbb{R}^n$ is a column vector.

Exercise 6.4.13. Let $X = \begin{pmatrix} 0 & -u \\ u^t & 0 \end{pmatrix} \in \mathfrak{m}$ and $\tilde{X}$ the vector field on M associated to X. Let $v \in T_{p_0}M$. Show that

(1) $\tilde{X}_{p_0} = ru$;
(2) $\nabla_v \tilde{X} = 0$.

We want now to find an explicit expression for $\langle R(x,y)z,w\rangle$, where $x,y,z,w \in \mathbb{R}^n = T_{p_0}M$. Let $\tilde{X}$ (respectively,$\tilde{Y}, \tilde{Z}, \tilde{W}$) be the vector field derived as above from $X = \begin{pmatrix} 0 & -x \\ x^t & 0 \end{pmatrix}$ (respectively, from the matrices that result after replacing y, z, w for x). Using that $\nabla_v \tilde{A} = 0$ for all $v \in T_{p_0}M$ and all $A \in \mathfrak{m}$, we have:

$$r^4 \langle R_{p_0}(x,y)z,w\rangle = \langle R_{p_0}(\tilde{X},\tilde{Y})\tilde{Z},\tilde{W}\rangle = \tilde{X}_{p_0}\langle \nabla_{\tilde{Y}}\tilde{Z},\tilde{W}\rangle - \tilde{Y}_{p_0}\langle \nabla_{\tilde{X}}\tilde{Z},\tilde{W}\rangle. \tag{6.2}$$

The first term of the right hand side can be developed as follows (at the last line we use that $YXp_0 = -\langle y,x\rangle p_0$, with similar expressions for the other vectors):

$$\begin{aligned} \tilde{X}_{p_0}\langle \nabla_{\tilde{Y}}\tilde{Z},\tilde{W}\rangle &= \frac{d}{dt}|_{t=0}\langle \nabla_{\tilde{Y}}\tilde{Z},\tilde{W}\rangle_{e^{-tX}p_0} \\ &= -\frac{d}{dt}|_{t=0}\langle ZYe^{-tX}p_0, We^{-tX}p_0\rangle \\ &= \langle ZYXp_0, Wp_0\rangle + \langle ZYp_0, WXp_0\rangle \\ &= -\langle YXp_0, ZWp_0\rangle + \langle ZYp_0, WXp_0\rangle \\ &= -\langle y,x\rangle\langle z,w\rangle r^2 + \langle z,y\rangle\langle w,x\rangle r^2. \end{aligned}$$

We can now express the curvature tensor as follows.

$$\langle R_{p_0}(x,y)z,w\rangle = (-\langle y,x\rangle\langle z,w\rangle + \langle z,y\rangle\langle w,x\rangle - \langle x,y\rangle\langle z,w\rangle + \langle z,x\rangle\langle w,y\rangle)r^{-2}.$$

In particular, if x, y are orthogonal unit vectors and V is the plane spanned by them,

$$K(V) = \langle R(x,y)y,x\rangle = \frac{1}{r^2}.$$

5. The Laplacian

Let M be a Riemannian manifold with Riemannian metric $\langle\cdot,\cdot\rangle$.

6. *Grad.* If $f : M \to \mathbb{R}$ is a differentiable function, then $u \in T_pM \mapsto uf \in \mathbb{R}$ is a linear functional on T_pM. Therefore there exists a unique vector, $\mathrm{grad}_p f$, such that

$$uf = \langle \mathrm{grad}_p f, u\rangle.$$

If $e_1, \dots, e_n$ is an orthonormal basis for T_pM, then it follows that

$$\mathrm{grad}_p f = \sum_{i=1}^{n} (e_i f) e_i.$$

7. *Div.* Let ∇ be the Levi-Civita connection of M. If X is a vector field on M then

$$u \in T_pM \mapsto \nabla_u X \in T_pM$$

is a linear map. Define

$$\mathrm{div}_p X := \mathrm{Trace}(u \mapsto \nabla_u X).$$

Notice that for any choice of orthonormal basis $\{e_1, \dots, e_n\}$ of T_pM, we have

$$\mathrm{div}_p X = \sum_{i=1}^{n} \langle e_i, \nabla_{e_i} X \rangle_p.$$

For example, on $\mathbb{R}^n$, with the Euclidian connection D, the divergence of a vector field $X = \sum_{i=1}^n h_i e_i$, where $e_i = \frac{\partial}{\partial x_i}$, is

$$\mathrm{div}_p X = \sum_{i=1}^{n} \langle \frac{\partial}{\partial x_i}, \nabla_{\frac{\partial}{\partial x_i}} X \rangle = \sum_{i=1}^{n} \frac{\partial h_i}{\partial x_i}.$$

EXERCISE 6.5.1. Let f be a scalar function and X a vector field on a Riemannian manifold M. Define the linear map $df \otimes X : u \in T_pM \mapsto (uf)X_p \in T_pM$. Note that the trace of this map is given, with respect to an orthonormal basis, by $\sum_{i=1}^n (e_i f)\langle e_i, X \rangle$. Show that

$$\mathrm{div}_p(fX) = \mathrm{Trace}(df \otimes X) + f(p)\mathrm{div}_p X.$$

8. *Laplacian.* Let f be a twice differentiable function on a Riemannian manifold M. We define the Laplacian of f by

$$\Delta f := \mathrm{div}(\mathrm{grad} f).$$

The given definition of the Laplacian has the following equivalent formulation. Define the *Hessian* of f (relative to a given choice of connection ∇, which for us will always be the Levi-Civita connection on the Riemannian M), to be the linear map $H_p f : T_pM \to T_pM$ such that

$$(H_p f)u := \nabla_u \mathrm{grad} f.$$

Then the Laplacian of f at a point p can also be defined as the trace of the Hessian of f at p.

EXERCISE 6.5.2. Check that $\nabla_p f = \mathrm{Trace}(H_p f)$.

EXERCISE 6.5.3. Check that $\mathrm{Hess}_p f(u, v) := \langle (H_p f)u, v \rangle$ is a symmetric bilinear form.

It is interesting to point out that if p is a critical point of f, that is, $uf = 0$ for all $u \in T_pM$, then the Hessian of f at p does not depend on the choice of connection. We could more generally have defined the Hessian as the bilinear form $(u, v) \mapsto \mathrm{Hess}_p f(u, v) := (\nabla_u df)(v)$, where differentiation is applied to the 1-form df rather than its dual vector field $\mathrm{grad} f$. With this definition, the Hessian of f clearly does not depend on a metric at all, but only on a general connection ∇. Then $\mathrm{Hess}_p f$ and $H_p f$ are related by metric duality, that is $\mathrm{Hess}_p f(u, v) = \langle (H_p f)u, v \rangle$.

If, moreover, p is a critical point of f, then $\mathrm{Hess}_p f(u,v) = u(Vf) = v(Uf)$, where U, V are arbitrary differentiable local extensions of u, v.

EXERCISE 6.5.4. Show that the Hessian has the following property:

$$\frac{1}{2}\mathrm{Hess}_p(f^2) = f\mathrm{Hess}_p f + df \otimes df,$$

where $df \otimes df$ is the bilinear form defined by

$$(df \otimes df)(u,v) = (uf)(vf).$$

We could have defined the Hessian as an operation on function that associates to each differentiable f and $p \in M$ a symmetric bilinear form that satisfies the property of Exercise 6.5.4. It would then be possible to show that such an abstract Hessian uniquely defines a connection for which $\mathrm{Hess} f$ is the bilinear form ∇df.

9. *Expressing Δ in a local o.n. frame.* Let $\{Z_1, \dots, Z_n\}$ be a local orthonormal frame of vector fields on a Riemannian manifold M. Then

$$\begin{aligned}
\mathrm{div}(\mathrm{grad} f) &= \sum_{i=1}^{n} \left\langle Z_i, \nabla_{Z_i}\left(\sum_{j=1}^{n} (Z_j f) Z_j\right)\right\rangle \\
&= \sum_{i,j=1}^{n} \langle Z_i, (Z_i Z_j f) Z_j + (Z_j f)\nabla_{Z_i} Z_j\rangle \\
&= \sum_{i=1}^{n} Z_i^2 f + \sum_{i,j=1}^{n} \langle Z_i, \nabla_{Z_i} Z_j\rangle Z_j f \\
&= \sum_{i=1}^{n} Z_i^2 f - \sum_{i,j=1}^{n} \langle Z_j, \nabla_{Z_i} Z_i\rangle Z_j f.
\end{aligned}$$

EXERCISE 6.5.5. Let G be a Lie group with a bi-invariant metric $\langle \cdot, \cdot \rangle$, show that

$$\Delta f = \sum_{i=1}^{n} Z_i^2 f,$$

where $\{Z_1, \dots, Z_n\}$ is any orthonormal frame of left-invariant vector fields.

EXERCISE 6.5.6 (Conformally flat metrics). Let $U \subset \mathbb{R}^2$ be an open set and $h : U \to (0, \infty)$ a differentiable function. Define a metric on U by

$$\langle \cdot, \cdot \rangle := \frac{1}{h^2} \langle \cdot, \cdot \rangle^{\mathrm{euc}}.$$

Show that

$$\Delta f = h^2 \Delta^{\mathrm{euc}} f.$$

EXERCISE 6.5.7 (Surfaces of Revolution). Let $M \subset \mathbb{R}^3$ be a surface of revolution, spanned by rotating around an axis the graph of the function $t \mapsto R(t)$. Notice that a parametrization of M is given by the map

$$\varphi(t, \theta) = (R(t)\cos\theta, R(t)\sin\theta, t).$$

Show that

$$\Delta f = \frac{1}{R^2}\frac{\partial^2 f}{\partial \theta^2} + \frac{\partial^2 f}{\partial t^2}.$$

EXERCISE 6.5.8 (Laplacian on Hypersurfaces). Let M be a hypersurface in $\mathbb{R}^{n+1}$. Denote by Δ the Laplacian on M and by Δ^{euc} the Laplacian on $\mathbb{R}^{n+1}$. Let N be a unit vector field on some open set $U \subset \mathbb{R}^{n+1}$ containing M such that $N(p)$ is perpendicular to T_pM at each $p \in M$. Given a (twice) differentiable function f on M, let $\tilde{f}$ be any (twice) differentiable extension of f to U. Show that

$$(\Delta f)_p = (\Delta^{\text{euc}}\tilde{f})_p + (\nabla_N^{\text{euc}}N)_p\tilde{f} - N_p(N\tilde{f}),$$

for each $p \in M$.

EXERCISE 6.5.9. Use the previous exercise to show that if M is the sphere of radius R with center at the origin, then

$$(\Delta f)_p = (\Delta^{\text{euc}}\tilde{f})_p - \frac{1}{R}\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial \tilde{f}}{\partial r}\,(rp/R)\right)_{r=R}$$

for any extension $\tilde{f}$ of f. In particular, if we extend f so as to be linear along rays issuing from the origin, then $(\Delta f)_p = (\Delta^{\text{euc}}\tilde{f})_p$.

6. The Exponential Map and Normal Coordinates

Let M be a Riemannian manifold. Recall that a geodesic that passes through p at time $t = 0$ with velocity $v \in T_pM$ is a curve $\gamma(t)$ that satisfies the second order differential equation $\frac{\nabla\gamma'(t)}{dt} = 0$ with initial conditions $\gamma(0) = p$, $\gamma'(0) = v$. This initial value problem aways has a unique solution on a sufficiently small interval $-a \leq t \leq a$. For simplicity, we will assume that the Riemannian metric on M is *complete*, that is, geodesics exist for all time. It is well known that this is equivalent to completeness of M as a metric space for the metric defined as the shortest length among curves joinning two points.

The geodesic starting at p with tangent vector $v \in T_pM$ will be denoted by $\gamma(t) = \exp_p(tv)$. Taking $t = 1$, we obtain a map

$$v \in T_pM \mapsto \exp_p v \in M,$$

which is called the *exponential map* at p. It is not difficult to show that $\exp_p$ is a smooth map for each p. Moreover,

$$\begin{aligned} d(\exp_p)_0 v &= \frac{d}{dt}|_{t=0}\exp_p(tv) \\ &= \gamma'(0) \\ &= v. \end{aligned}$$

(We have above identified T_pM with its tangent space at 0.) It follows from the Inverse Function Theorem that on some sufficiently small neighborhood of p, the exponential has a smooth inverse. Therefore we can use $\exp_p$ to produce local parametrizations of M.

Let now $\sigma : \mathbb{R}^n \to T_pM$ be a linear isometry. Then the local parametrization

$$u \in \mathbb{R}^n \mapsto \exp_p \sigma u \in M$$

defines a *normal coordinate system* centered at p. We will adopt the notation $\exp_\sigma u := \exp_p \sigma u$. (Here, σ is to be regarded as an orthonormal frame at p.)

EXERCISE 6.6.1. Let G be a matrix Lie group with bi-invariant metric. This is the case, for example, if G is a compact, connected subgroup of $GL(m,\mathbb{R})$, for some m. Let $Z_1,\dots,Z_n$ be an orthonormal basis for the Lie algebra $\mathfrak{g}=T_eG$ of G. We are regarding the elements of $\mathfrak{g}$ as $m\times m$ matrices. Set

$$\sigma : u\in\mathbb{R}^n \mapsto u_1Z_1+\cdots+u_nZ_n\in\mathfrak{g}.$$

Show that

$$\exp_e \sigma u = e^{u_1Z_1+\cdots+u_nZ_n}.$$

(On the right hand side we have ordinary matrix exponential.)

a. The Metric in Normal Coordinates. Let $\varphi : W\subset\mathbb{R}^n\to U\subset M$ be the parametrization associated to a normal coordinate system at $p\subset U$. The coordinate vector fields will be denoted by $X_i=\varphi_*\frac{\partial}{\partial x_i}$. A vector field on U of the form $X=a_1X_1+\cdots+a_nX_n$, where $a_1,\dots,a_n$ are constants, will be called a *constant vector field* (relative to given normal coordinates).

Let $G(x):=G_{ij}(x):=\langle X_i,X_j\rangle_{\varphi(x)}$. It turns out the the coefficients in the Taylor series expansion of $G(x)$ are functions of the curvature tensor and its covariat derivatives.

We can write the Taylor approximation of $G(x)$ as follows:

$$G(x)=G(0)+G_1(x)+\frac{1}{2}G_2(x)+\cdots+\frac{1}{n!}G_n(x)+\frac{1}{(n+1)!}\frac{d^{n+1}}{dt^{n+1}}|_{t=t^*}G(tx)$$

for some $t^*\in(0,1)$, where $G_k(x)=\frac{d^k}{dt^k}|_{t=0}G(tx)$ is a homogeneous polynomial in the components of x, whose coefficients we wish to calculate.

We would like now to describe the first few terms of this expansion. This will require the following remarks, which we state as exercises.

EXERCISE 6.6.2. Writing $G_k(x)=\sum_{i_1,\dots,i_k=1}^n a_{i_1\dots i_k}x_{i_1}\dots x_{i_k}$, show that

$$a_{i_1\dots i_k}=\frac{\partial^k G_{ij}}{\partial x_{i_1}\dots\partial x_{i_k}}(0)=(X_{i_1}\dots X_{i_k}\langle X_i,X_j\rangle)_p.$$

EXERCISE 6.6.3. If X is any constant vector field relative to a normal coordinate system at p, then show that $\nabla_uX=0$ for all $u\in T_pM$. (By polarization it suffices to show that $(\nabla_XX)_p=0$ for all constant vector fields. But this is immediate from the fact that the flow line of X through p is a geodesic.) Use this to show that

$$G(x)=\delta_{ij}+\frac{1}{2}\sum_{k,l=1}^n a_{kl}x_kx_l+\dots.$$

Notice that, if h is a smooth function and Y_i are constant vector fields for some normal coordinate system, then $Q(Y_1,\dots,Y_k):=Y_1\dots Y_kh$ is a k-linear symmetric form. Therefore, once we know how to calculate $(X^kG)(p)$, for a general constant vector field X, then all terms $(X_{i_1}\dots X_{i_k}\langle X_i,X_j\rangle)_p$ follow by polarization (or linearization). In fact, writing $X=s_1X_1+\cdots+s_nX_n$, we have

$$a_{i_1\dots i_k}=\frac{\partial^k(X^kG)_p}{\partial s_{i_1}\dots\partial s_{i_k}}|_{s=0}.$$

To tackle the second order terms in the Taylor series of G, we thus need to regard the function $X^2\langle X_i,X_j\rangle$.

First, note that $(\nabla_X\dots\nabla_XX)_q=0$ whenever q is a point on the geodesic through p with velocity X_p. By expanding the equation $\nabla_{X+sY}\nabla_{X+sY}(X+sY)=0$

in powers of s and equating the coefficients to 0 we obtain (from the linear term in s)

$$\nabla_X\nabla_X Y + \nabla_X\nabla_Y X + \nabla_Y\nabla_X X = 0.$$

On the other hand, since $[X, Y] = 0$ (whenever X, Y are constant vector fields for the given normal coordinate system), we have

$$\nabla_X\nabla_X Y = \nabla_X\nabla_Y X = R(X, Y)X + \nabla_Y\nabla_X X.$$

It follows that

$$\nabla_Y\nabla_X X = -2\nabla_X\nabla_X Y = -2\nabla_X\nabla_Y X = -\frac{2}{3}R(X, Y)X.$$

Another polarization, this time by expanding

$$\nabla_{X+sZ}\nabla_{X+sZ}Y = \frac{1}{3}R(X + sZ, Y)(X + sZ)$$

as a polynomial in s, yields:

$$\nabla_X\nabla_Z Y + \nabla_Z\nabla_X Y = \frac{1}{3}(R(Z, Y)X + R(X, Y)Z). \tag{6.1}$$

Before moving on with the calculation, we state in the next exercise the main algebraic identities for the curvature tensor. The last expression is known as the Bianchi identity. It holds for any torsion-free connection.

Exercise 6.6.4. Prove the following identities for the curvature tensor associated to the Levi-Civita connection of an arbitrary Riemannain metric.

$$\langle R(X, Y)Z, T\rangle = -\langle R(Y, X)Z, T\rangle \tag{6.2}$$

$$\langle R(X, Y)Z, T\rangle = -\langle R(X, Y)T, Z\rangle \tag{6.3}$$

$$\langle R(X, Y)Z, T\rangle = \langle R(Z, T)X, Y\rangle \tag{6.4}$$

$$R(X, Y)Z = R(X, Z)Y - R(Y, Z)X. \tag{6.5}$$

With the above identities, the fact that $\nabla_u X = 0$ for all $u \in T_pM$ when X is a constant vector field, and by the expression 6.1 we obtain, after some algebraic manipulation, (Bianchi's identity is not used):

$$\begin{aligned}(XY\langle Z, T\rangle)_p &= \langle\nabla_X\nabla_Y Z, T\rangle_p + \langle Z, \nabla_X\nabla_Y T\rangle_p \\ &= \frac{1}{6}\left(\langle R(Y, Z)X + R(X, Z)Y, T\rangle\right).\end{aligned}$$

We can now write the Taylor polynomial of order two of the metric coefficients. The remaining manipulations leading to it are left as an exercise.

Exercise 6.6.5. Set $R_{ijkl} = \langle R(X_i, X_j)X_k, X_l\rangle_p$. Then

$$\langle X_i, X_j\rangle_{\varphi(x)} = \delta_{ij} + \frac{1}{6}\sum_{a_1\leq a_2}(R_{a_1ia_2j} + R_{a_1ja_2i})x_{a_1}x_{a_2} + \text{error term of order 3.}$$

Exercise 6.6.6. The $(M, \langle\cdot,\cdot\rangle)$ is said to be *locally symmetric* if $\nabla R = 0$. This means that the tensor

$$T(X, Y, Z, T, V) := \langle(\nabla_X R)(Y, Z)T, V\rangle$$

is identically zero. Show that if the Riemannian manifold is locally symmetric, then the terms of order 3 in the previous exercise are 0.

7. The Orthonormal Frame Bundle

Let M be an n-dimensional Riemanninan manifold with metric $\langle \cdot, \cdot \rangle$. Associated to each tangent space T_pM is the space $O_p(M)$ of all orthogonal frames (basis of T_pM) at p. It will be convenient to regard a frame at p as a linear isomorphism $\sigma : \mathbb{R}^n \to T_pM$, from which a basis is obtained by taking $\{\sigma e_1, \ldots, \sigma e_n\}$, $\{e_1, \ldots, e_n\}$ being the standard basis of $\mathbb{R}^n$. The frame is orthonormal if the basis $\{\sigma e_1, \ldots, \sigma e_n\}$ is orthonormal.

Let $O(M) = \cup_{p\in M} O_p(M)$, the set of all orthonormal frames at all points of M. We denote by $\pi : O(M) \to M$ the base point map: $\pi(\sigma) = p$ exactly when $\sigma \in O_p(M)$.

Exercise 6.7.1. $O(M)$ has a unique smooth manifold structure for which $\pi : O(M) \to M$ is a smooth map and such that the following holds. If $X_1, \ldots, X_n$ are smooth vector fields on some open set $U \subset M$ such that $X_1(p), \ldots, X_n(p)$ is an orthonormal basis of T_pM at each $p \in U$, then

$$\sigma : p \in U \mapsto (a \in \mathbb{R}^n \mapsto \sigma_p a = a_1 X_1(p) + \ldots a_n X_n(p) \in T_pM)$$

is a smooth map from U to $O(M)$ such that $\pi \circ \sigma$ is the identity on U. With this structure, the map $U \times O(n) \to \pi^{-1}(U) \subset O(M)$ defined by $\varphi : (p, A) \mapsto \sigma_p \circ A$ is a diffeomorphism such that $\varphi(p, AB) = \varphi(p, A)B$.

The main convenience in expressing frames as linear maps from $\mathbb{R}^n$ to T_pM is in expressing the (right-) action of the group $O(n)$ on $O(M)$. In fact, given $\sigma \in O_p(M)$ and $A \in O(n)$ (the latter viewed as a linear map from $\mathbb{R}^n$ to $\mathbb{R}^n$), then $(\sigma, A) \mapsto \sigma \circ A$ is another element of $O_p(M)$.

Exercise 6.7.2. Show that the action of $O(n)$ on $O_p(M)$ is a smooth action whose orbits are precisely the fibers $O_p(M)$ and the action on each fiber is simply transitively. In fact, fixing $\sigma \in O_p(M)$, the map $A \in O(n) \mapsto \sigma \circ A \in O_p(M)$ is a diffeomorphism.

We will also use the notation $\mathcal{O}^M := O(M)$. The fiber above p will be written $\mathcal{O}^p$, whereas $\mathcal{O}^U := \pi^{-1}(U)$.

The fibers $\mathbb{O}^p$ have a distinguished set of vector fields, obtained as follows. Let X be any left-invariant vector field on $O(n)$. The 1-parameter subgroup of $O(n)$ generated by X will be written $t \mapsto e^{tX}$. Then X gives rise to a vector field $\tilde{X}$ on $\mathcal{O}^p$ by setting

$$\tilde{X}_\sigma := \frac{d}{dt}|_{t=0} \sigma \circ e^{tX} =: \sigma X.$$

Notice that $\tilde{X}$ is not a right-invariant vector field on $\mathcal{O}^M$, unless $n = 2$ (in which case $O(n)$ is abelian.) In fact, $(L_A)_* \tilde{X} = \widetilde{A^{-1} X A}$

Denote by $V_\sigma = T_\sigma \mathcal{O}^p$. This is a subspace of $T_\sigma \mathcal{O}^M$ having the same dimension as the Lie algebra of $O(n)$, called the *vertical subspace* at σ. Since the kernel of $d\pi_\sigma : T_\sigma \mathcal{O} \to T_pM$ is precisely V_σ, the quotient space $T_\sigma \mathcal{O}^M / V_\sigma$ is a linear isomorphism.

It is natural to ask whether there should exist a complementary subspace H_σ for V_σ. We now show that a Riemannian connection can be interpreted precisely as such a *horizontal distribution* of subspaces of $T\mathcal{O}^M$. The subspace H_σ will be given as the image of an injective map $K_\sigma : T_pM \to T_\sigma \mathcal{O}^M$ such that $d\pi_\sigma \circ K_\sigma$ is the identity map on T_pM. K_σ is defined as follows. Let $v \in T_pM$ and $\gamma(t)$ any differentiable curve such that $\gamma(0) = p$ and $\gamma'(0) = v$. Let $\xi(t)$ denote the parallel translation of σ along $\gamma(t)$, with respect to the Levi-Civita connection. Notice that $\xi(t)$ is a differentiable curve in $\mathcal{O}^M$ such that $\xi(0) = \sigma$ and $\pi(\xi(t)) = \gamma(t)$. We set

$$K_\sigma v := \xi'(0).$$

It is immediate that

$$d\pi_\sigma K_\sigma v = \frac{d}{dt}|_{t=0}\pi \circ \xi = \gamma'(0) = v.$$

With $H_\sigma := K_\sigma T_p M$, we now have that $T_\sigma \mathcal{O}^M = V_\sigma \oplus H_\sigma$.

A very nice fact about $\mathcal{O}^M$ is that this is a parallelizable manifold. In other words, there is on $\mathcal{O}^M$ a globally defined, smooth frame of vector fields. This can be seen as follows. Let $e_1, \dots, e_n$ be the standard basis of $\mathbb{R}^n$ and $X_1, \dots, X_N$ a basis for the Lie algebra of $O(n)$ (so that $N = n(n-1)/2$). Let $\tilde{X}_i$ be the vertical vector fields defined earlier and set

$$Z_i(\sigma) := K_\sigma(\sigma e_i).$$

Then Z_i are smooth horizontal vector fields and we have a frame

$$\tilde{X}_1, \dots, \tilde{X}_N, Z_1, \dots, Z_n,$$

which is adapted to the decomposition of $T\mathcal{O}^M$ into vertical and horizontal subbundles.

Note that if $\xi(t)$, $\xi(0) = \sigma$, is the parallel translation of a frame σ along a curve $\gamma(t)$, then $\xi(t)A$ is the parallel translation of the frame $\sigma \circ A$ along the same curve, for each invertible linear map $A : \mathbb{R}^n \to \mathbb{R}^n$. The claim in the next exercise is an immediate consequence of this fact.

Exercise 6.7.3. Show that the horizontal subbundle H is preserved by the right-action of $O(n)$ on $\mathcal{O}^M$. In other words, let $R_A(\sigma) := \sigma \circ A$, for $A \in O(n)$. Then

$$(dR_A)_\sigma H_\sigma = H_{R_A(\sigma)}$$

for all $\sigma \in \mathcal{O}^M$ and $A \in O(n)$.

Exercise 6.7.4. Show that parallel transport has now the following description. Let $w \in T_p M$ and $\gamma(t)$ a differentiable curve on M such that $\gamma(0) = p$. Choose a frame $\sigma \in \mathcal{O}^p$ and let $w_0 := \sigma^{-1} w \in \mathbb{R}^n$. Now let $\xi(t)$ be the unique horizontal lift of $\gamma(t)$ that starts at σ. In other words, $\xi(t)$ is a differentiable curve in $\mathcal{O}^M$ starting at σ such that $\xi'(t) \in H_{\xi t}$ for all t and $\pi \circ \xi = \sigma$. Then

$$t \mapsto \xi(t) w_0 = \xi(t)\xi(0)^{-1} w$$

is the parallel translation of w along $\gamma(t)$. Show that $\xi(t)$ is indeed unique and that parallel translation does not depend on the choice of σ.

It will be convenient to use the following notations: given $X \in \mathfrak{o}(n)$, we write the corresponding vertical vector field as either $\mathcal{V}(X)$ or $\mathcal{V}^X$. If $w \in \mathbb{R}^n$, we define the horizontal vector field $\sigma \mapsto \mathcal{H}_\sigma(w) := \mathcal{H}^w_\sigma := K_\sigma \sigma w$.

Exercise 6.7.5. Let $\gamma(t)$ be a geodesic with initial velocity v at $p \in M$. Let $\sigma \in \mathcal{O}^p$ and denote by $\sigma(t)$ the parallel transport of σ along $\gamma(t)$. Finally, let $f : M \to \mathbb{R}$ be a smooth function. Show that

$$\mathcal{H}^v_\sigma \mathcal{H}^v (f \circ \pi) = \left.\frac{d^2}{dt^2}\right|_{t=0} f(\gamma(t)).$$

EXERCISE 6.7.6. Using the previous exercise, show that the Riemannian Laplacian, Δ, on M has the following characterization. Choose an arbitrary orthonormal basis $e_1, \dots, e_n$ for $\mathbb{R}^n$, let $Z_i := \mathcal{H}^{e_i}$ be the horizontal vector field corresponding to e_i, and set $\Delta^{\mathcal{O}} = \sum_{i=1}^n Z_i^2$. Show that $\Delta = \pi_* \Delta^{\mathcal{O}}$. In other words, if $f : M \to \mathbb{R}$ is a smooth function on M, then

$$(\Delta f) \circ \pi = \Delta^{\mathcal{O}}(f \circ \pi).$$

We give now a somewhat different description of the horizontal subbundle, in terms of a $\mathfrak{o}(n)$-valued 1-form on $\mathcal{O}$.

Let $\xi \in T_\sigma \mathcal{O}^M$. Then ξ is the velocity vector of some curve, $\sigma(t)$, such that $\sigma(0) = \sigma$. Since $\sigma(t)$ is a frame field along $\gamma(t) := \pi(\sigma(t))$, it makes sense to consider its covariant derivative $\frac{\nabla}{dt}\big|_{t=0} \sigma(t)$. (The covariant derivative of the frame is, by definition, the assignment, for each $x \in \mathcal{R}^n$, of the covariant derivative of the vector field $\sigma(t)x$ along $\gamma(t)$.) The result, $\frac{\nabla}{dt}\big|_{t=0} \sigma(t)$, is a linear map (possibly singular) from $\mathbb{R}^n$ to $T_p M$, so it also makes sense to define a linear map from $\mathbb{R}^n$ to $\mathbb{R}^n$ by the composition

$$\Theta_\sigma(\xi) := \sigma(0)^{-1} \circ \left(\frac{\nabla}{dt}\bigg|_{t=0} \sigma(t) \right).$$

EXERCISE 6.7.7. Show that $\Theta_\sigma(\xi)$ is skew-symmetric; in other words, it is an element of the Lie algebra of the group $O(n)$.

Therefore, Θ can be regarded as a (matrix valued) 1-form on the manifold $\mathcal{O}^M$, taking values in the Lie algebra of the orthogonal group. This is the so called *connection form*.

We discuss next some of the properties of Θ.

EXERCISE 6.7.8. Let $X \in \mathfrak{o}(n)$ and $\tilde{X}$ the corresponding vertical vector field. Show that $\Theta(\tilde{X}) = X$.

EXERCISE 6.7.9. Show that the null-space of Θ is precisely the horizontal subbundle.

Before stating the next exercise, we make the following remarks. Let

$$\sigma : U \to \mathcal{O}^U$$

be a smooth section, and denote $(Z_i)_p := \sigma(p)e_i$ the orthonormal basis at p associated to the frame field σ. Let $\bar{Z}_i$ denote the unique tangent vector field to the submanifold $\sigma(U)$ (of the frame bundle of M) that projects to Z_i under the base point map $\pi : \mathcal{O}^M \to M$. Also let Θ_{rs} be the (r, s)-entry of the matrix Θ.

EXERCISE 6.7.10. Show the following:

(1) Θ_{rs} is a 1-form on $\sigma(U)$ such that $\Theta_{rs}(\bar{Z}_i) = \langle \nabla_{Z_i} Z_r, Z_s \rangle \circ \pi$;
(2) $\bar{Z}_j \Theta_{rs}(\bar{Z}_i) = \langle \nabla_{Z_j} \nabla_{Z_i} Z_r, Z_s \rangle \circ \pi + \langle \nabla_{Z_i} Z_r, \nabla_{Z_j} Z_s \rangle \circ \pi$;
(3) Using $d\Theta_{rs}(\bar{Z}_i, \bar{Z}_j = \bar{Z}_i \Theta_{rs}(\bar{Z}_j) - \bar{Z}_j \Theta_{rs}(\bar{Z}_i) - \Theta_{rs}([\bar{Z}_i, \bar{Z}_j])$, show that

$$d\Theta_{rs}(\bar{Z}_i, \bar{Z}_j) = \langle R(Z_i, Z_j) Z_r, Z_s \rangle + \sum_{k=1}^n \left(\Theta_{rk}(\bar{Z}_i) \Theta_{ks}(\bar{Z}_j) - \Theta_{rk}(\bar{Z}_j) \Theta_{ks}(\bar{Z}_i) \right).$$

The last expression in the exercise can be written more compactly as

$$(d\Theta - \Theta \wedge \Theta)_\sigma(\bar{Z}_i, \bar{Z}_j) = \sigma^{-1} R(Z_i, Z_j) \sigma.$$

The *curvature form* is the 2-form Ω on $\mathcal{O}^M$ with values in $\mathfrak{o}(n)$ defined by

$$\Omega(X, Y) = (d\Theta - \Theta \wedge \Theta)_\sigma(X^H, Y^H),$$

where X^H and Y^H denote the horizontal components of X and Y, respectively.

CHAPTER 7

Differential Forms and Integration

1. Differential Forms

a. Determinants. Recall that $M(n,\mathbb{R})$ denotes the set of n by n real matrices and that, as a vector space, $M(n,\mathbb{R})$ is isomorphic to the direct sum of $\mathbb{R}^n$ with itself n times. A matrix $A \in M(n,\mathbb{R})$ may be written as an n-tuple $A = (A_1, \dots, A_n)$, where $A_i \in \mathbb{R}^n$ is a column vector.

We define the *determinant* function $D : M(n,\mathbb{R}) \to \mathbb{R}$ (which will often be written as det later in the notes) by the following properties.

1. $A_i \mapsto D(A_1, \dots, A_i, \dots, A_n)$ is a linear map for each i. (So D is an n-linear function from $\mathbb{R}^n \times \cdots \times \mathbb{R}^n$ into $\mathbb{R}$.)
2. For any two indices $i \neq j$, the sign of D changes by interchanging A_i and A_j: $D(A_1, \cdots, A_i, \cdots, A_j, \cdots, A_n) = -D(A_1, \cdots, A_j, \cdots, A_i, \cdots, A_n)$. In particular, the determinant is zero whenever two different columns are scalar multiples of one another.
3. $D(e_1, \dots, e_n) = 1$, i.e., if I is the identity matrix we define $D(I) = 1$.

Let S_n be the group of permutations of $\{1, \dots, n\}$. For each $\sigma \in S_n$, the symbol $(-1)^\sigma$ will denote the sign of the permutation. Since any element of S_n can be written as a product of transpositions, it follows from properties 2 and 1 that $D(e_{\sigma(1)}, \dots, e_{\sigma(1)}) = (-1)^\sigma$.

Exercise 7.1.1. Show that D satisfying the above three properties exists, is unique, and is given by the expression

$$D(A) = \sum_{\sigma \in S_n} (-1)^\sigma a_{1\sigma(1)} \cdots a_{n\sigma(n)}.$$

In particular, D is a polynomial in the entries of A.

The next proposition should already be known to you.

Proposition 7.1.2. Let A be a matrix in $M(n,\mathbb{R})$. Then the following are equivalent.

1. $D(A) \neq 0$.
2. There does not exist a nonzero (column) vector $v \in \mathbb{R}^n$ such that $Av = 0$.
3. There is a matrix A^{-1} (the *inverse of* A) such that $AA^{-1} = A^{-1}A = I$.

Proof. The proof is elementary (and standard) and is left to you as an exercise. □

Lemma 7.1.3. The set of matrices A for which $D(A) \neq 0$ is open and dense in $M(n,\mathbb{R})$.

Proof. The set is open since D is a continuous function (a polynomial, in fact). Suppose that it is not dense, so that we can find A such that $D(B) = 0$

for all B close enough to A (in the natural topology of $M(n,\mathbb{R})$ that makes the identification with $\mathbb{R}^{n^2}$ a homeomorphism). Consider the function

$$R(h_1,\ldots,h_n) := D(A_1+h_1,\ldots,A_n+h_n).$$

$R(h_1,\ldots,h_i,\ldots,h_n)$ is a polynomial in the entries of the h_i that vanishes identically for all $(h_1,\ldots,h_n)$ such that $\sum_{i=1}^n |h_i|^2$ is sufficiently small. Therefore R must vanish identically. (If a polynomial vanishes identically on a neighborhood of 0, it must be the zero polynomial. Why?) Therefore D must vanish identically, a contradiction. □

PROPOSITION 7.1.4. $D(AB) = D(A)D(B)$.

PROOF. If we show the identity assuming that $D(A) \neq 0$, then the proposition will hold for all A, due to the lemma and since D continuous. Thus suppose $D(A) \neq 0$ and define $U(B) := D(AB)/D(A)$. Note that $AB = (AB_1,\ldots,AB_n)$, where each column vector AB_i is the matrix product of A by B_i. Written this way it becomes clear (and is left to you to verify) that $U(B)$ also satisfies the properties defining D. Consequently, by the uniqueness of D claimed in the first exercise, we must have $U(B) = D(B)$ and the proposition holds. □

It follows from the previous proposition that the subset $GL(n,\mathbb{R}) \subset M(n,\mathbb{R})$ consisting of *nonsingular matrices*, i.e. matrices with nonzero determinant, forms a group under the operation of matrix multiplication. This group is called the *general linear group*. The subset of $GL(n,\mathbb{R})$ consisting of matrices of determinant 1 is a subgroup, called the *special linear group*. It is denoted $SL(n,\mathbb{R})$.

PROPOSITION 7.1.5. $D(A^t) = D(A)$ for each $A \in M(n,\mathbb{R})$.

PROOF. For each $\sigma \in S_n$ we have $(-1)^\sigma = (-1)^{\sigma^{-1}}$ and

$$\begin{aligned} a_{1\sigma(1)}\cdots a_{n\sigma(n)} &= a_{\sigma^{-1}(\sigma(1))\sigma(1)}\cdots a_{\sigma^{-1}(\sigma(n))\sigma(n)} \\ &= a_{\sigma^{-1}(1)1}\cdots a_{\sigma^{-1}(n)n}. \end{aligned}$$

(The last line involved rearranging the terms in the product.) Therefore,

$$\begin{aligned} D(A) &= \sum_{\sigma\in S_n} (-1)^{\sigma^{-1}} a_{\sigma^{-1}(1)1}\cdots a_{\sigma^{-1}(n)n} \\ &= \sum_{\sigma\in S_n} (-1)^{\sigma} a_{\sigma(1)1}\cdots a_{\sigma(n)n} \\ &= D(A^t). \end{aligned}$$

□

Let $\langle u,v\rangle$ denote the ordinary dot product of vectors u and v in $\mathbb{R}^n$. If $u = \sum_{i=1}^n a_i e_i$ and $v = \sum_{i=1}^n b_i e_i$ then, by definition,

$$\langle u,v\rangle = \sum_{i=1}^n a_i b_i.$$

Note that $|u| = \langle u,u\rangle^{1/2}$.

$A \in M(n,\mathbb{R})$ is said to be an *orthogonal matrix* if $|Au| = |u|$ for all $u \in \mathbb{R}^n$. This is equivalent to

$$\langle Au, Av\rangle = \langle u,v\rangle$$

for all $u, v \in \mathbb{R}^n$. (To see the equivalence, expand $|A(u+v)|^2 = |u+v|^2$ and compare the terms.)

LEMMA 7.1.6. $\langle A^t u, v\rangle = \langle u, Av\rangle$ for all $u, v \in \mathbb{R}^n$.

PROOF. By linearity, it suffices to show the identity for the basis elements $u = e_i$ and $v = e_j$. It is then immediate that $\langle A^t e_i, e_j\rangle = a_{ij} = \langle e_i, Ae_j\rangle$, where a_{ij} is the (i, j)-entry of A. □

PROPOSITION 7.1.7. $A \in M_n(n)$ is orthogonal if and only if $A^t = A^{-1}$.

PROOF. Note that $\langle u, v\rangle = \langle Au, Av\rangle = \langle A^t Au, v\rangle$ for all u and v. Therefore $A^t A = I$, the identity matrix. □

The subset of $M(n, \mathbb{R})$ consisting of orthogonal matrices is denoted $O(n)$. It is clear that $O(n)$ is a subgroup of $GL(n, \mathbb{R})$. Moreover, the previous proposition shows that the determinant of an orthogonal matrix is either 1 or -1. The subgroup consisting of orthogonal matrices of determinant 1 is denoted $SO(n)$, and is called the *special orthogonal group*. Note that if $v_1, \dots, v_n$ form an orthonormal basis of $\mathbb{R}^n$ (that is, $\langle v_i, v_j\rangle = 0$ for $i \neq j$ and it is 1 if $i = j$), then the matrix $A = (v_1, \dots, v_n)$ belongs to $O(n)$. Conversely, the column vectors of an orthogonal matrix form an orthonormal basis of $\mathbb{R}^n$.

We say that an ordered set of vectors $u_1, \dots, u_n$ in $\mathbb{R}^n$ is a *positive basis* of $\mathbb{R}^n$ if the determinant of the matrix $(u_1, \dots, u_n)$ is positive. Otherwise (if the determinant is negative) we say that it is a *negative basis*.

EXERCISE 7.1.8. Show that $O(n)$ is a compact subset of $M(n, \mathbb{R})$. (It is clearly a closed subset. The main point is to show that it is also bounded.)

The next exercise gives an interpretation of the Gram-Schmidt orthogonalization process. $P(n, \mathbb{R})$ will denote the group of all upper-triangular real matrices with positive diagonal entries. (Check that this is indeed a group.)

EXERCISE 7.1.9. Show that any matrix $A \in GL(n, \mathbb{R})$ can be written uniquely as a product $A = RU$, where $R \in O(n)$ and $T \in P(n, \mathbb{R})$. More precisely, show that there exists a continuous bijection

$$GL(n, \mathbb{R}) \to O(n) \times P(n, \mathbb{R})$$

whose inverse is $(R, U) \mapsto RU$. Use this fact to conclude that $GL(n, \mathbb{R})$ is homeomorphic to $O(n) \times \mathbb{R}^{n(n+1)/2}$.

Recall that λ is an *eigenvalue* of $A \in M(n, \mathbb{R})$ if $D(A - \lambda I) = 0$. A nonzero vector $v \in \mathbb{R}^n$ is an *eigenvector* associated to λ if $Av = \lambda v$. If A is a *symmetric matrix*, which means that $A^t = A$, then $\mathbb{R}^n$ admits an orthonormal basis $\{v_1, \dots, v_n\}$ consisting of eigenvectors of A. This is a standard fact from linear algebra. A non-standard proof of this fact using Lagrange multipliers will be suggested later in an exercise. It follows from this that any symmetric matrix A can be written as a product KLK^{-1}, where $K \in O(n)$ and L is diagonal, with diagonal entries given by the eigenvalues of A.

Note that the eigenvalues of $A^t A$ are all nonnegative. In fact, if $A^t Av = \lambda v$, we have $\langle Av, Av\rangle = \langle A^t Av, v\rangle = \lambda\langle v, v\rangle$ so that $\lambda = \langle Av, Av\rangle/\langle v, v\rangle \geq 0$.

LEMMA 7.1.10. For each $A \in M(n, \mathbb{R})$ there is $K \in O(n)$ and a diagonal matrix L with nonnegative entries such that $A^t A = (KLK^{-1})^2$.

PROOF. Since A^tA is symmetric, we can write $A^tA = KDK^{-1}$, where K is orthogonal and D is diagonal with nonnegative entries. Therefore, $D = L^2$, and the lemma follows. □

The matrix KLK^{-1} obtained in the previous proposition will be denoted $(A^tA)^{1/2}$. Suppose now that A is nonsingular and define $K = A(A^tA)^{-1/2}$. It is immediate that $K \in O(n)$. Therefore, a nonsingular square matrix A can always be written in the form $A = KS$ where $K \in O(n)$ and S is symmetric. Using again the fact that a symmetric matrix admits an orthonormal basis of eigenvectors, there must exist an orthogonal matrix K_1 and a diagonal matrix L_1 such that $S = K_1^{-1}L_1K_1$. Therefore, any nonsingular square matrix can be written in the form K_0D1K_1, where D is diagonal and K_0, K_1 are orthogonal.

PROPOSITION 7.1.11. Any matrix $A \in M(n, \mathbb{R})$ can be written as $A = K_1DK_2$, where $K_1, K_2 \in O(n)$ and D is a diagonal matrix with nonnegative entries.

PROOF. This has just been shown for A nonsingular. Any singular A is the limit of a sequence of nonsingular matrices $A(n) = K_1(n)L(n)K_2(n)$. Since $O(n)$ is compact, we can pass to a subsequence such that $K_i(n)$ converges to K_i. Therefore $L(n) = K_1(n)^{-1}A(n)K_2(n)^{-1}$ converges (for that subsequence) to a nonnegative diagonal matrix D. It follows that $A = K_1DK_2$. □

Given $A \in M(n, \mathbb{R})$, let P_A denote the parallelepiped in $\mathbb{R}^n$ spanned by the vectors $A_1, \dots, A_n$. This is the set of all $v = t_1A_1 + \cdots + t_nA_n$ for $t_i \in [0, 1]$. (See figure 2.)

PROPOSITION 7.1.12. The volume of the parallelepiped P_A is equal to $|D(A)|$.

PROOF. First note that P_A is the image under the linear transformation A of the standard parallelepiped spanned by the basis vectors $e_1, \dots, e_n$, which has volume 1. On the other hand, elements of $O(n)$ do not change lengths and angles since, $\langle Au, Av\rangle = \langle u, v\rangle$ (recall that the dot product determines the notion of angle by the cosine formula $\langle u, v\rangle = |u||v| \cos\theta$, where θ is the angle between u and v) so A also does not change volumes. Therefore, if $A = K_1LK_2$, with L diagonal, the volume of P_A is the same as the volume of P_L. But the volume of P_L is easily seen to be, up to sign, the determinant of L, which is just the product of the diagonal entries. (We are using here the definition of the volume of a product of intervals as the product of the lengths of those intervals.) Moreover, $D(K_1LK_2) = D(K_1)D(L)D(K_2) = \pm D(L)$, since the determinant of an orthogonal matrix is either 1 or -1. The proposition now follows. □

THEOREM 7.1.13. Let E be a solid region in $\mathbb{R}^n$ of finite volume. Let AE be the image of E under a linear transformation $A : \mathbb{R}^n \to \mathbb{R}^n$. Then

$$\mathrm{Vol}(AE) = |D(A)|\mathrm{Vol}(E).$$

PROOF. It is been implicitly assumed here that the volume of E can be obtained as a limit of the sum of the volumes of small cubes whose union approximates E. Since by the previous proposition the volume of each cube transforms under A in the way claimed by the theorem, the volume of E will transform in the same way. □

EXERCISE 7.1.14. Show that by identifying $M(n,\mathbb{R})$ with $\mathbb{R}^{n^2}$ in a natural way, the dot product of $\mathbb{R}^{n^2}$ corresponds to $\langle X, Y\rangle = \mathrm{Tr}(X^tY)$, $X, Y \in M(n,\mathbb{R})$. Use this to find the gradient of the determinant function.

EXERCISE 7.1.15. If D is the determinant function, show that

$$dD_AV = D(A)\mathrm{Tr}(VA^{-1}).$$

If V and W are two vector spaces, the set of all linear transformations from V to W is also a vector space, denoted $\mathrm{Hom}(V,W)$. Note that if V has dimension n and W has dimension m, then $\mathrm{Hom}(V,W)$ is isomorphic to the space of m by n matrices. The *dual space* of V, denoted V^*, is by definition the n-dimensional space $\mathrm{Hom}(V,\mathbb{R})$. If $v_1, \ldots, v_n$ form a basis of V, the *dual basis* of V^* consists of the vectors v_i^* such that $v_i^*(v_j)$ is 0 if $i \neq j$ and 1 if $i = j$.

The space $\mathrm{Hom}(V,V)$ is also denoted $\mathrm{End}(V)$. If $L \in \mathrm{End}(V)$, the *trace* of L is defined by

$$\mathrm{Tr}(L) = \sum_{i=1}^{n} v_i^*(Lv_i).$$

EXERCISE 7.1.16. Show that the definition of $\mathrm{Tr}(L)$ does not depend on the choice of basis for V. Also show that if $L \in \mathrm{End}(\mathbb{R}^n) = M(n,\mathbb{R})$, then $\mathrm{Tr}(L)$ is given by the sum of the diagonal entries.

2. Differential forms in $\mathbb{R}^n$

Recall that $(e_1, \ldots, e_n)$ denotes the natural basis of $\mathbb{R}^n$. As before, the dual basis will be denoted $(dx_1, \ldots, dx_n)$, so that $dx_i(e_j)$ is 0 when $i \neq j$ and 1 if $i = j$. Note that if we use the derivative notation for the basis vectors, i.e. $e_i = \frac{\partial}{\partial x_i}$, then

$$dx_i(e_j) = dx_i(\frac{\partial}{\partial x_j}) = \frac{\partial x_i}{\partial x_j} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

where we have used the definition $v_pf = df_pv$ of v_p. From now on, we will give preference to the derivative notation for vectors. It should be kept in mind that from this point of view vectors become attached to points in $\mathbb{R}^n$, since a vector is a first order derivative *at* a certain point. The set of vectors at a point p will be denoted $\mathbb{R}^n_p$. The vector space $\mathbb{R}^n_p$, which we sometimes denote by $T\mathbb{R}^n_p$ will be referred to as the *tangent space* to $\mathbb{R}^n$ at p.

A differential 1-*form* in $\mathbb{R}^n$ is a field of dual vectors depending differentiably on the point. More precisely, α is a differential 1-form on $\mathbb{R}^n$ if

$$\alpha = f_1dx_1 + \cdots + f_ndx_n$$

where $f_1, \ldots, f_n$ are differentiable functions on $\mathbb{R}^n$. Note that a 1-form defines a map from $\mathbb{R}^n$ into $(\mathbb{R}^n)^*$

For example, if $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth function, its differential

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i$$

is a differential 1-form.

A *k-form* in $\mathbb{R}^n$ is a map that associates to each $p \in \mathbb{R}^n$ an *alternating* k-linear map

$$\varphi_p : \mathbb{R}^n_p \times \cdots \times \mathbb{R}^n_p \to \mathbb{R}.$$

Alternating means that for any vectors $v_1, \dots, v_k \in \mathbb{R}^n_p$

$$\varphi_p(v_{\sigma(1)}, \dots, v_{\sigma(k)}) = (-1)^\sigma \varphi_p(v_1, \dots, v_k)$$

for any permutation σ of k symbols.

For example, the determinant function may be regarded as an n-form on $\mathbb{R}^n$. More generally, given 1-forms $\varphi_1, \dots, \varphi_k$, we obtain a k-form by defining

$$(\varphi_1 \wedge \cdots \wedge \varphi_k)(v_1, \dots, v_k) := \det(\varphi_i(v_j)).$$

It follows from the properties of determinants that $\varphi_1 \wedge \cdots \wedge \varphi_k$ is k-linear and alternating.

Note that

$$\varphi_{\sigma(1)} \wedge \cdots \wedge \varphi_{\sigma(k)} = (-1)^\sigma \varphi_1 \wedge \cdots \wedge \varphi_k$$

for any permutation σ of k numbers. In particular, $\varphi_1 \wedge \cdots \wedge \varphi_k = 0$ whenever there are two indices $i \neq j$ and constant c such that $\varphi_i = c\varphi_j$.

If $I = (i_1, \dots, i_k)$ is an ordered set with $i_l \in \{1, \dots, n\}$, we define

$$dx_I := dx_{i_1} \wedge \cdots \wedge dx_{i_k}.$$

EXERCISE 7.2.1. If $I = (1, 2, \dots, n)$ and $A = (A_1, \dots, A_n)$ is a matrix whose columns A_i are regarded as vectors in $\mathbb{R}^n$, show that

$$(dx_I)_p(A_1, \dots, A_n) = \det(A).$$

EXERCISE 7.2.2. If $f = (f_1, \dots, f_n)$ is a differentiable map from $\mathbb{R}^n$ to $\mathbb{R}^n$, show that

$$(df_1)_p \wedge \cdots \wedge (df_n)_p = \det(Jf(p)) dx_1 \wedge \cdots \wedge dx_n.$$

The space of alternating k-linear forms on $\mathbb{R}^n_p$ will be denoted $\Lambda^k(\mathbb{R}^n)^*$.

PROPOSITION 7.2.3. $\Lambda^k(\mathbb{R}^n_p)^*$ is a vector space and It has a basis given by The set

$$\{(dx_I)_p \mid I = (i_1, \dots, i_k), 1 \leq i_1 < \cdots < i_k \leq n\}$$

is a basis. In particular,

$$\dim \Lambda^k(\mathbb{R}^n)^* = \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

PROOF. It is clear that $\Lambda^k(\mathbb{R}^n)^*$ is a vector space. The claim about dimension will follow once we show that the set given is indeed a basis. The key observation is that if $I = (i_1, \dots, i_k)$, $J = (j_1, \dots, j_k)$, with $i_1 < \cdots < i_k$ and $j_1 < \cdots < j_k$, then

$$dx_I(e_{j_1}, \dots, e_{j_k}) = \begin{cases} 0 & \text{if } \{i_1, \dots, i_k\} \neq \{j_1, \dots, j_k\} \\ 1 & \text{if } \{i_1, \dots, i_k\} = \{j_1, \dots, j_k\} \end{cases}$$

as a simple calculation shows. We can show that the dx_I are linearly independent as follows. If

$$\varphi = \sum_{i_1 < \cdots < i_k} a_{i_1 \cdots i_k} dx_{i_1} \wedge \cdots \wedge dx_{i_k} = 0$$

then

$$\varphi(e_{j_1}, \dots, e_{j_k}) = a_{j_1 \cdots j_k} = 0.$$

It remains to show that the dx_I span $\Lambda(\mathbb{R}^n_p)^*$. Let $\varphi \in \Lambda(\mathbb{R}^n_p)^*$ and define

$$\omega := \sum_{i_1 < \cdots < i_k} \varphi(e_{i_1}, \dots, e_{i_k}) dx_{i_1} \wedge \cdots \wedge dx_{i_k}.$$

It follows that

$$\omega(e_{i_1}, \ldots, e_{i_k}) = \varphi(e_{i_1}, \ldots, e_{i_k})$$

for any k-tuple of basis vectors. By k-linearity we must have $\omega = \varphi$. $\square$

Since the space of k-linear maps in $\mathbb{R}^n_p = \mathbb{R}^n$ is a vector space, it makes sense to define C^k maps from $\mathbb{R}^n$ into it. A C^k *exterior k-form* in $\mathbb{R}^n$ is a C^k function from $\mathbb{R}^n$ into that vector space. If φ is an exterior k-form, we can write by the previous proposition

$$\varphi_p = \sum_{i_1 < \cdots < i_k} a_{i_1 \ldots i_k}(p) dx_{i_1} \wedge \cdots \wedge dx_{i_k} = \sum_I a_I(p) dx_I.$$

The coefficients a_I are functions of $\mathbb{R}^n$. These functions are C^k if and only if φ is a C^k exterior k-form.

We will regard functions as 0-forms. The space $\Lambda^0(\mathbb{R}^n)^*$ is simply $\mathbb{R}$.

The space of exterior k-forms is a vector space under the operations of addition

$$\sum_I a_I dx_I + \sum_I b_I dx_I = \sum_I (a_I + b_I) dx_I$$

multiplication by a constant

$$c \sum_I a_I dx_I = \sum_I c a_I dx_I.$$

If $\omega = \sum_I a_I dx_I$ is an exterior k-form on $\mathbb{R}^n$ and f is a real valued function on $\mathbb{R}^n$, then $f\omega = \sum_I f a_I dx_I$ is also an exterior k-form. With these operations, the space of smooth exterior k-forms is an algebra over the ring of smooth functions.

The *exterior product*, or *wedge product*, of a k-form $\omega = \sum_I a_I dx_I$ and l-form $\theta = \sum_I b_I dx_I$ is the $k+l$-form defined by

$$\omega \wedge \theta = \sum_{I,J} a_I b_J dx_I \wedge dx_J.$$

For example, given the 1-form $\omega = x_2 dx_1 - x_1 x_4^2 dx_3 + x_3 dx_4$ and the 2-form $\theta = \cos x_3 dx_2 \wedge dx_4$, their exterior product is the 3-form

$$\omega \wedge \theta = x_2 \cos x_3 dx_1 \wedge dx_2 \wedge dx_4 + x_1 x_4 \cos x_3 dx_2 \wedge dx_3 \wedge dx_4.$$

EXERCISE 7.2.4. If

$$\begin{aligned} f_1(\rho, \theta, \varphi) &= \rho \sin\varphi \cos\theta \\ f_2(\rho, \theta, \varphi) &= \rho \sin\varphi \sin\theta \\ f_3(\rho, \theta, \varphi) &= \rho \cos\varphi, \end{aligned}$$

show that $df_1 \wedge df_2 \wedge df_3 = \rho^2 \sin\varphi d\rho \wedge d\theta \wedge d\varphi$.

PROPOSITION 7.2.5. Let ω be a k-form, φ an s-form, and θ an r-form. Then:

(1) $(\omega \wedge \varphi) \wedge \theta = \omega \wedge (\varphi \wedge \theta)$
(2) $\omega \wedge (\varphi + \theta) = \omega \wedge \varphi + \omega \wedge \theta$ (for $r = s$)
(3) $\omega \wedge \varphi = (-1)^{ks} \varphi \wedge \omega$.

PROOF. Only 3 requires some work. The key remark is that given $I = (i_1, \ldots, i_k)$ and $J = (j_1, \ldots, j_s)$, with $i_1 < \cdots < i_k$ and $j_1 < \cdots < j_s$, then

$$dx_I \wedge dx_J = (-1)^{ks} dx_J \wedge dx_I.$$

But this is clear since ks is precisely the number of transpositions of the type $dx_i \wedge dx_j = -dx_j \wedge dx_i$ that are needed to get dx_I entirely past dx_J. The details are left for the reader. □

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a differentiable map. Then f induces a map f^* that carries k-forms on $\mathbb{R}^m$ to k-forms on $\mathbb{R}^n$, defined as follows. Let ω be a k-form on $\mathbb{R}^m$, $p \in \mathbb{R}^n$, and $v_1, \ldots, v_n$ n vectors in $\mathbb{R}^n_p$. Note that the differential df_p sends vectors in $\mathbb{R}^n_p$ to vectors in $\mathbb{R}^m_{f(p)}$. We define $f^*\omega$ by

$$(f^*\omega)_p(v_1, \ldots, v_n) = \omega_{f(p)}(df_p v_1, \ldots, df_p v_n).$$

We regard functions $g : \mathbb{R}^n \to \mathbb{R}$ as 0-forms. The induced action of f on a 0-form is defined by $f^*g = g \circ f$.

Exercise 7.2.6. Show that if $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}$ are differentiable, then the chain rule can be expressed as

$$d(f^*g) = f^*dg$$

where the differentials $d(f^*g)$ and dg are regarded as 1-forms.

The form $f^*\omega$ will be called the *pull-back* of ω under f. The next proposition gives some of the properties of the pull-back.

Proposition 7.2.7. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a differentiable map, let $g : \mathbb{R}^m \to \mathbb{R}$ be a 0-form on $\mathbb{R}^m$, let ω and φ be k-forms on $\mathbb{R}^m$ and let ν be an s-form on $\mathbb{R}^m$. Then:

(1) $f^*(\omega + \varphi) = f^*\omega + f^*\varphi$
(2) $f^*(g\omega) = f^*g f^*\omega$
(3) $f^*(\omega \wedge \nu) = f^*\omega \wedge f^*\nu$.

Proof. The proof of 1 and 2 are immediate and are left to the reader. Note that 2 is a special case of 3. We prove here property 3. Using properties 1 and 2, it suffices to show that if $\varphi_1, \ldots, \varphi_k$ are 1-forms, then

$$f^*(\varphi_1 \wedge \cdots \wedge \varphi_k) = f^*\varphi_1 \wedge \cdots \wedge f^*\varphi_k.$$

To verify this last claim, let $v_1, \ldots, v_k$ be vectors at p. Then

$$\begin{aligned} f^*(\varphi_1 \wedge \cdots \wedge \varphi_k)(v_1, \ldots, v_k) &= (\varphi_1 \wedge \cdots \wedge \varphi_k)_{f(p)}(df_p v_1, \ldots, df_p v_k) \\ &= \det(\varphi_i(df_p v_j)) \\ &= \det((f^*\varphi_i) v_j) \\ &= (f^*\varphi_1 \wedge \cdots \wedge f^*\varphi_k)(v_1, \ldots, v_k). \end{aligned}$$

□

Exercise 7.2.8. If $f : \mathbb{R}^n \to \mathbb{R}^n$ is differentiable and $I = (1, \ldots, n)$, show that

$$(f^*dx_I)_p = \det(Jf(p))(dx_I)_p.$$

Exercise 7.2.9. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable and $x_1, \ldots, x_m$ be the coordinates of $\mathbb{R}^m$. If $dx_I = dx_{i_1} \wedge \cdots \wedge dx_{i_k}$, show that

$$f^*dx_I = df_{i_1} \wedge \cdots \wedge df_{i_k}.$$

Note that exercise 1.5.4 can be interpreted as follows. If $f = (f_1, f_2, f_3)$ is the function defined there (the coordinate change to spherical coordinates) and x, y, z are used instead of x_1, x_2, x_3, then

$$f^*(dx \wedge dy \wedge dz) = \rho^2 \sin\varphi d\rho \wedge d\theta \wedge d\varphi.$$

PROPOSITION 7.2.10. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^l \to \mathbb{R}^n$ be differentiable maps. Then*

$$(f \circ g)^* = g^* \circ f^*.$$

PROOF. Let ω be a k-form and $v_1, \ldots, v_k \in \mathbb{R}^l_p$. Then

$$\begin{aligned} ((f \circ g)^*\omega)_p(v_1, \ldots, v_k) &= \omega_{(f\circ g)(p)}(d(f \circ g)_p v_1, \ldots, d(f \circ g)_p v_k) \\ &= \omega_{f(g(p))}(df_{g(p)} dg_p v_1, \ldots, df_{g(p)} dg_p v_k) \\ &= (f^*\omega)_{g(p)}(dg_p v_1, \ldots, dg_p v_k) \\ &= (g^*(f^*\omega))_p(v_1, \ldots, v_k). \end{aligned}$$

□

We now define a derivation on form that extends the differential df of a function (0-form). Let $\omega = \sum_I a_I dx_I$ be a differential k-form on $\mathbb{R}^n$. The *exterior derivative* of ω is defined by

$$d\omega = \sum_I da_I \wedge dx_I.$$

For example, if $\omega = z^2 dx + \sin x dy + xy dz$ is a 1-form in $\mathbb{R}^3$, then

$$\begin{aligned} d\omega &= d(z^2) \wedge dx + d(\sin x) \wedge dy + d(xy) \wedge dz \\ &= 2z dz \wedge dx + \cos x dx \wedge dy + (y dx + x dy) \wedge dz \\ &= -2z dx \wedge dz + \cos x dx \wedge dy + y dx \wedge dz + x dy \wedge dz \\ &= \cos x dx \wedge dy + (y - 2z) dx \wedge dz + x dy \wedge dz. \end{aligned}$$

EXERCISE 7.2.11. Let $\omega = f dx + g dy + h dz$ be a 1-form on $\mathbb{R}^3$, where f, g, h are differentiable functions. Find $d\omega$ and show that $dd\omega = 0$. Also show that if $f : \mathbb{R}^m \to \mathbb{R}$ is a C^2 function (0-form), then $ddf = 0$.

It is clear that the exterior derivative is linear: if ω_1 and ω_2 are k-forms and a, b are constants, then

$$d(a\omega_1 + b\omega_2) = a d\omega_1 + b d\omega_2.$$

It is also a simple calculation to check that if h is a differentiable function and ω is a k-form, then

$$d(h\omega) = dh \wedge \omega + h d\omega.$$

Further properties of the exterior derivative are enumerated in the next proposition.

PROPOSITION 7.2.12. *Let ω be a k-form and φ an s-form, both on $\mathbb{R}^m$. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a differentiable map. Then,*

(1) $d(\omega \wedge \varphi) = d\omega \wedge \varphi + (-1)^k \omega \wedge d\varphi$
(2) $d^2\omega := dd\omega = 0$
(3) $d(f^*\omega) = d(f^*\omega)$

PROOF. In all these properties, it is sufficient to assume that $\omega = adx_I$ and $\varphi = bdx_I$, where a and b are functions since the general cases would follow immediately by linearity of the operations involved. Proposition 1.5.5 implies that if ν is any 1-form,

$$\nu \wedge dx_I = (-1)^k dx_I \wedge \nu.$$

Therefore

$$\begin{aligned} d(abdx_I \wedge dx_J) &= d(ab) \wedge dx_I \wedge dx_J \\ &= (bda + adb) \wedge dx_I \wedge dx_J \\ &= da \wedge dx_I \wedge (bdx_J) + adb \wedge dx_I \wedge dx_J \\ &= da \wedge dx_I \wedge (bdx_J) + (-1)^k (adx_I) \wedge db \wedge dx_J \\ &= d(\omega) \wedge \varphi + (-1)^k \omega \wedge d\varphi \end{aligned}$$

showing 1. To prove 2, note that

$$d^2(adx_I) = d(da \wedge dx_I) = d^2 a \wedge dx_I + (-1)^1 da \wedge d(dx_I) = d^2 a \wedge dx_I.$$

But $d^2 a = 0$ by exercise 1.5.11.

To prove 3, first note that by properites 1 and 2,

$$d(f^* dx_I) = d(df_{i_1} \wedge \cdots \wedge df_{i_k}) = 0.$$

Moreover, by exercise 1.5.6,

$$d(f^* a) = f^* da.$$

Combining these two properties we get:

$$\begin{aligned} d(f^*(adx_I)) &= d(f^* a f^* dx_I) \\ &= d(f^* a) \wedge f^* dx_I \\ &= (f^* da) \wedge f^* dx_I \\ &= f^*(da \wedge dx_I) \\ &= f^* d(adx_I). \end{aligned}$$

□

EXERCISE 7.2.13. Let ω be the 2-form on $\mathbb{R}^{2n}$ given by

$$\omega = dx_1 \wedge dx_2 + \cdots + dx_{2n-1} \wedge dx_{dx}.$$

Show the following:

(1) $\omega^n = dx_1 \wedge dx_2 \wedge \cdots \wedge dx_{2n-1} \wedge dx_{2n}$.
(2) If $A : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is a linear transformation (defined by a $2n$ by $2n$ matrix A) such that $A^*\omega = \omega$, then $\det A = 1$.
(3) The set of all $A \in M(2n, \mathbb{R})$ such that $A^*\omega = \omega$ forms a group. (It is called the *real symplectic group*, and is often denoted $Sp(2n, \mathbb{R})$.)

3. Differential Forms on Manifolds

Let M be a smooth manifold. The space of alternating k-linear maps

$$\omega : T_pM \times \cdots \times T_pM \to \mathbb{R}$$

will be denoted $\bigwedge^k (T_pM)^*$. Let $\bigwedge^k (TM)^*$ be the disjoint union of the $\bigwedge^k (T_pM)^*$, for $p \in M$. We denote by $\pi : \bigwedge^k (TM)^* \to M$ the base point map.

The set $\bigwedge^k(TM)^*$ can be given a manifold structure as follows. Let $\varphi : U_0 \subset \mathbb{R}^n \to U \subset M$ be a smooth local parametrization of M. Let $\{dx_I\}$ be the standard basis of k-forms on $\mathbb{R}^n$ and $\alpha_I = (\varphi^{-1})^* dx_I$ the alternating k-form such that for each $p \in U$ and $u_1, \ldots, u_k \in T_pM$, gives

$$\alpha_I(u_1, \ldots, u_k) := dx_I(d\varphi_k^{-1}u_1, \ldots, d\varphi_p^{-1}u_k).$$

Now define a local parametrization of $\bigwedge^k(TM)^*$ by the assignment

$$\begin{aligned}\phi : U_0 \times \mathbb{R}^{\binom{n}{k}} &\to \pi^{-1}(U)\\ (q, \{a_I\}) &\mapsto \sum_I a_I \alpha_I(\varphi(q)).\end{aligned}$$

The notation is chosen so that the change of coordinates corresponds to the (already defined) pull-back map for alternating forms on $\mathbb{R}^n$.

Exercise 7.3.1. Show that the given parametrizations actually define a smooth manifold structure on $\bigwedge^k(TM)^*$ of dimension $n + \binom{n}{k}$.

A *differential k-form* on M is defined as a smooth map (section) $\omega : M \to \bigwedge^k(TM)^*$ such that $\pi \circ \omega$ is the identity map on M.

The definitions that were given earlier for forms on $\mathbb{R}^n$ can be immediately transferred to forms on M. Take, for example, the wedge product. Let ω be a k-form and ν an l-form on M. Define $\omega \wedge \nu$ as the $k + l$-form such that

$$(\omega \wedge \nu) + p := \left[(\varphi^{-1})^*((\varphi^*\omega) \wedge (\varphi^*\nu))\right]_p$$

for p in the image of a parametrization $\varphi : U_0 \to U$.

Exercise 7.3.2. Show that the definition of $\wedge$ does not depend on the choice of parametrization. (It is used here that $f^*(\xi \wedge \eta) = f^*\xi \wedge f^*\eta$ holds for forms on $\mathbb{R}^n$.)

Similarly, one defines the exterior derivative, $d\omega$, on M by means of the exterior derivative on $\mathbb{R}^n$. If φ is a local parametrization of M and ω is a k-form on M, then for all p in the image of φ we set:

$$(d\omega)_p = ({\varphi^{-1}}^*(d\varphi^*\omega))_p.$$

Exercise 7.3.3. Show that $d\omega$ does not depend on the choice of local parametrization. (It is used here that for forms on $\mathbb{R}^n$, d commutes with pull-back.)

Exercise 7.3.4. With the definitions given above, show that if $F : M \to N$ is a smooth map between two manifolds and ω, ν are forms on N, then

(1) $F^*(\omega \wedge \nu) = F^*\omega \wedge F^*\nu$, and
(2) $dF^*\omega = F^*d\omega$.

4. Lie Derivative

The Lie derivative of a k-form ω on M with respect to a vector field X is defined by

$$\mathcal{L}_X\omega := \left.\frac{d}{dt}\right|_{t=0} \Phi_t^*\omega$$

where Φ_t denotes the flow of X.

EXERCISE 7.4.1. Show that if ω is a one form and X, Y are vector fields on M, then

$$X\omega(Y) = (\mathcal{L}_X\omega)(Y) + \omega(\mathcal{L}_X Y).$$

It should be recalled that $\mathcal{L}_X Y = [X, Y]$.

EXERCISE 7.4.2. Let ω and ν be differential forms on M, and X a smooth vector field. Show that

$$\mathcal{L}_X(\omega \wedge \nu) = (\mathcal{L}_X\omega) \wedge \nu + \omega \wedge \mathcal{L}_X\nu.$$

If ω is a $k+1$-form and X is a vector field on M, then $i_X\omega$ indicates the k-form that, to a set of vectors $u_1, \ldots, u_k$, gives

$$(i_X\omega)(u_1, \ldots, u_k) = \omega(X, u_1, \ldots, u_k).$$

EXERCISE 7.4.3. Show that $\mathcal{L}_X\omega = i_X d\omega + di_X\omega$.

If ω is a 1-form and X, Y are vector fields, then it follows from the identities contained in the previous exercises that

$$\begin{aligned} d\omega(X, Y) &= (i_X d\omega)(Y) \\ &= (\mathcal{L}_X\omega)(Y) - (di_X\omega)(Y) \\ &= X\omega(Y) - \omega([X, Y]) - Y\omega(X). \end{aligned}$$

Therefore,

$$d\omega(X, Y) = X\omega(Y) - Y\omega(X) - \omega([X, Y]).$$

This useful expression generalizes as follows.

EXERCISE 7.4.4. Let ω be a k-form on M and $X_1, \ldots, X_{k+1}$ smooth vector fields. Show that $d\omega(X_1, \ldots, X_{k+1}) = (I) + (II)$, where

$$\begin{aligned} (I) &= \sum_{i=1}^{k+1} (-1)^{i+1} X_i\omega(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{k+1}) \\ (II) &= \sum_{i<j} (-1)^{i+j} \omega([X_i, X_j], X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{j-1}, X_{j+1}, \ldots, X_{k+1}). \end{aligned}$$

5. The Laplacian on forms

We introduce now some further derivative operations on functions and forms in $\mathbb{R}^n$ that will make use of the geometry of $\mathbb{R}^n$. Distances and angles are encoded in the definition of the dot product, which we have been and will continue to denote by $\langle \cdot, \cdot \rangle$.

Fix a positive orthonormal basis $(u_1, \ldots, u_n)$. (Recall that the basis is positive if the determinant of the matrix A that performs the change of basis from the given one to the standard basis $(e_1, \ldots, e_n)$ has positive determinant, equal to 1 if the first basis is orthonormal.) Each u_i determines by duality via the dot product a 1-form $\varphi_i = \langle u_i, \cdot \rangle$ on $\mathbb{R}^n$. The k-forms $\varphi_I := \varphi_{i_1} \wedge \cdots \wedge \varphi_{i_k}$, for $I = (i_1, \ldots, i_k)$ and $1 \leq i_1 < \cdots < i_k \leq n$, make up a basis of $\Lambda^k(\mathbb{R}^n)^*$.

We define a positive inner product on $\Lambda^k(\mathbb{R}^n)^*$ by declaring the basis consisting of the φ_I to be orthonormal. The inner product of k-forms will still be denoted by $\langle \cdot, \cdot \rangle$. Therefore, if $\omega = \sum_I a_I\varphi_I$ and $\theta = \sum_I b_I\varphi_I$, then

$$\langle \omega, \theta \rangle = \sum_I a_I b_I.$$

Note that if φ_I and φ_J are given by indices that are not necessarily ordered by increasing order, then

$$(*) \qquad \langle \varphi_I, \varphi_J \rangle = \det(R)$$

where R is the k by k matrix whose (r, s)-entry is $\langle u_{i_r}, u_{j_s} \rangle$.

This definition doe not depend on the choice of the orthonormal basis of $\mathbb{R}^n$, as will be seen in a moment. In particular, we can take $\varphi_I = dx_I$.

LEMMA 7.5.1. The definition of inner product of k-forms is independent of the choice of orthonormal basis.

PROOF. Let $v_1, \ldots, v_k$ and $w_1, \ldots, w_k$ be any set of $2k$ vectors. Define the 1-forms $\psi_i = \langle v_i, \cdot \rangle$ and $\eta_i = \langle w_i, \cdot \rangle$. Then, by expanding the vectors v_r, w_s in terms of the orthonormal basis, using $*$, above, and then recollecting the terms, we obtain

$$\langle \psi_1 \wedge \cdots \wedge \psi_k, \eta_1 \wedge \cdots \wedge \eta_k \rangle = \det(\langle v_i, w_j \rangle).$$

But this expression does not depend on the choice of basis, proving the claim. □

EXERCISE 7.5.2. Let $I = (i_1, \ldots, i_k)$ and $J = (j_1, \ldots, j_k)$ denote multi-indices such that $1 \leq i_1 < \cdots < i_k \leq n$ and $1 \leq j_1 < \cdots < j_k \leq n$. If $A = (a_{ij})$ is an n by n matrix, denote by $A|_{IJ}$ the k by k matrix whose (r, s)-entry is $a_{i_r j_s}$. Let δ_{IJ} be 0 if $I \neq J$ and 1 if $I = J$. Show that if A is an orthogonal matrix, the formula

$$\sum_L \det(A|_{IL}) \det(A|_{JL}) = \delta_{IJ}$$

holds, where the sum is over the set of $L = (l_1, \ldots, l_k)$ with $1 \leq l_1 < \cdots < l_k \leq n$. (Hint: The inner product on k-forms was defined so that the elements φ_I form an orthonormal basis. To show the above formula, use the fact that if ψ_I is obtained as φ_I from a second orthonormal basis, with change of basis matrix given by A, then the various ψ_I also form an orthonormal basis with respect to the same inner product.)

Let $(u_1, \ldots, u_n)$ be a positive orthonormal basis for $\mathbb{R}^n$, $\varphi_i := \langle u_i, \cdot \rangle$, and define the n-form $\nu := \varphi_1 \wedge \cdots \wedge \varphi_n$. It is clear that ν does not depend on the choice of positive orthonormal basis. (If ν' comes from another basis $(v_1, \ldots, v_n)$, then $\nu' = c\nu$, where c is the determinant of the matrix $R = (\langle v_i, u_j \rangle)$, which is the change of basis matrix. Assuming that the second basis is also positive orthonormal, R must be an orthogonal matrix of determinant 1.) We call ν the *volume form* of $\mathbb{R}^n$.

We saw before that $\Lambda^k(\mathbb{R}^n)^*$ has dimension $\binom{n}{k}$. Since $\binom{n}{k} = \binom{n}{n-k}$, we see that $\Lambda^k(\mathbb{R}^n)^*$ and $\Lambda^{n-k}(\mathbb{R}^n)^*$ have the same dimension, hence these two vector spaces are isomorphic. It turns out that the choice of inner product of $\mathbb{R}^n$ provides a concrete isomorphism between these two spaces.

LEMMA 7.5.3. Let $\theta \in \Lambda^k(\mathbb{R}^n)^*$. Then, there exists a unique $\star\theta \in \Lambda^{n-k}(\mathbb{R}^n)^*$ such that $\omega \wedge \star\theta = \langle \omega, \theta \rangle \nu$ for all $\omega \in \Lambda^k(\mathbb{R}^n)^*$. If $\theta = dx_{i_1} \wedge \cdots \wedge dx_{i_k}$, $i_1 < \cdots < i_k$, then

$$\star\theta = (-1)^\sigma dx_{j_1} \wedge \cdots \wedge dx_{j_{n-k}}$$

$(j_1 < \cdots < j_{n-k})$ where $\{j_1, \ldots, j_{n-k}\}$ is the complement of $\{i_1, \ldots, i_k\}$ in $\{1, \ldots, n\}$ and $(-1)^\sigma$ is the sign of the permutation

$$\sigma = \begin{pmatrix} i_1 & \ldots & i_k & j_1 & \ldots & j_{n-k} \\ 1 & \ldots & k & k+1 & \ldots & n \end{pmatrix}.$$

Moreover, $\star$ is a linear isomorphism between $\Lambda^k(\mathbb{R}^n)^*$ and $\Lambda^{n-k}(\mathbb{R}^n)^*$.

Proof. Uniqueness follows from the fact that if $\mu \in \Lambda^{n-k}(\mathbb{R}^n)^*$ and $\omega \wedge \mu = 0$ for all $\omega \in \Lambda^k(\mathbb{R}^n)^*$, then $\mu = 0$. (To check this it suffices to consider $\omega = dx_I$, where dx_I ranges over the standard basis of $\Lambda^k(\mathbb{R}^n)^*$.) Existence follows by checking that $\star dx_I = (-1)^\sigma dx_J$ satisfies $\omega \wedge \star dx_I = \langle \omega, dx_I \rangle \nu$ and extending $\star$ to $\Lambda^k(\mathbb{R}^n)^*$ by linearity. The details are left to the reader. That $\star$ is an isomorphism follows from the fact that the spaces of k and $n-k$-forms have the same dimension and that $\star\theta = 0$ implies that $\langle \theta, \theta \rangle = 0$, hence $\theta = 0$. □

The operation $\star$ is called the *Hodge star operator*.

Exercise 7.5.4. If $\omega = a_{12}dx_1 \wedge dx_2 + a_{13}dx_1 \wedge dx_3 + a_{23}dx_2 \wedge dx_3$ is a 2-form in $\mathbb{R}^3$, show that

$$\star\omega = a_{23}dx_1 - a_{13}dx_2 + a_{12}dx_3.$$

Exercise 7.5.5. If ω is a k-form in $\mathbb{R}^n$, show that $\star\star\,\omega = (-1)^{k(n-k)}\omega$.

We now show how the exterior derivative d can be combined with $\star$ to produce a number of interesting differential operators on forms and vector fields.

Let v be a smooth vector field on $\mathbb{R}^n$, and let $\omega = \langle v, \cdot \rangle$ be the smooth 1-form dual to v. Then $\star\omega$ is an $n-1$-form, so that $d \star \omega$ is an n-form. Therefore, there exists a smooth function $\operatorname{div}(v)$ such that $d \star \omega =: \operatorname{div}(v)\nu$. The function $\operatorname{div}(v)$ is called the *divergence* of v. Since $\star\nu = 1$, we can also write $\operatorname{div}(v) = \star d \star \omega$.

Exercise 7.5.6. Show that if $v = \sum_{i=1}^n f_i \frac{\partial}{\partial x_i}$, then $\operatorname{div}(v) = \sum_{i=1}^n \frac{\partial f_i}{\partial x_i}$.

If $n = 3$ and ω is, as before, the dual 1-form associated to v, then $\star d\omega$ is also a 1-form, hence there exists a (uniquely determined) vector field $\operatorname{curl}(v)$, called the *curl* of v, such that

$$\langle \operatorname{curl}(v), \cdot \rangle := \star d\omega.$$

Exercise 7.5.7. Show that if $v = P\frac{\partial}{\partial x_1} + Q\frac{\partial}{\partial x_2} + R\frac{\partial}{\partial x_3}$, then

$$\operatorname{curl}(v) = (\frac{\partial f_3}{\partial x_2} - \frac{\partial f_2}{\partial x_3})\frac{\partial}{\partial x_1} + (\frac{\partial f_1}{\partial x_3} - \frac{\partial f_3}{\partial x_1})\frac{\partial}{\partial x_2} + (\frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2})\frac{\partial}{\partial x_3}.$$

Exercise 7.5.8. Show that if v is a smooth vector field on $\mathbb{R}^3$, then $\operatorname{div}(\operatorname{curl}(v)) = 0$ and $\operatorname{curl}(\operatorname{grad}(h)) = 0$ are both consequences of the identity $dd = 0$.

The *codifferential* on forms is the differential operator defined bysuch that

$$\delta_{k+1}\omega := (-1)^{nk+1} \star d \star \omega$$

for any smooth $k+1$-form ω. The *Laplacian* on k-forms is defined by

$$\Delta_k := -(\delta_{k+1}d + d\delta_k).$$

We will often omit the subscript in δ_k and Δ_k.

Exercise 7.5.9. Show that

$$\Delta f = \delta df = \operatorname{div}(\operatorname{grad} f) = \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i^2}.$$

Exercise 7.5.10. Show that if ω is a smooth $k-1$-form and α is a smooth k-form, then $d(\omega \wedge \star\alpha) = \langle d\omega, \alpha \rangle \nu - \langle \omega, \delta\alpha \rangle \nu$.

6. Integration of Forms

Let ω be a differential form on M (or on an open set $U \subset M$). The *support* of ω is defined as the closure of the set of points $p \in M$ at which ω is not zero.

Suppose that $\omega = h(x_1, \ldots, x_n)dx_1 \wedge \cdots \wedge dx_n$ is a smooth n-form on an open set $U \subset \mathbb{R}^n$ whose support $K \subset U$ is compact. Define

$$\int_U \omega := \int_K h(x_1, \ldots, x_n)dx,$$

where the right-hand side is the ordinary integral of h on $\mathbb{R}^n$ with respect to the Lebesgue measure.

If ω is an n-form on an n-dimensional manifold M whose support K is contained in the image set of a smooth local parametrization, $\varphi : U_0 \to U$, we would like to define

$$\int_M \omega = \int_{\mathbb{R}^n} \varphi^*\omega$$

where $\varphi^*\omega = hdx_1 \wedge \cdots \wedge dx_n$.

There are a couple of issues to worry about at this point. First, does the definition depend on the choice of parametrization? (It does not, up to signe.) What to do if the support of ω is not contained in a single parametrized neighborhood? (We use a partition of unity to "localize" the integral to coordinate neighborhoods.) Implicit in these two questions is the issue of orientability of M.

Thus suppose that the support K of ω is contained in the intersection $U_1 \cap U_2$, where $U_i = \varphi_i(V_i)$ and $\varphi_i : V_i \to U_i$ is a smooth parametrization, $i = 1, 2$. Write $(\varphi_i)^*\omega = h_i dx_1 \wedge \cdots \wedge dx_n$ and $F = \varphi^{-1} \circ \varphi_2$. Notice that F is a diffeomorphism from $W_1 := \varphi_2^{-1}(U_1 \cap U_2)$ to $W_2 = \varphi_1(U_1 \cap U_2)$.

$$\begin{aligned} h_2(x)dx_1 \wedge \cdots \wedge dx_n &= (\varphi_2^*\omega)_x \\ &= (F^*\varphi_1^*\omega)_x \\ &= h_1(F(x))\det(dF_x)dx_1 \wedge \cdots \wedge x_n. \end{aligned}$$

On the other hand, we know from multivariable calculus that

$$\int_{W_1} h(F(x))|\det(dF_x)|dx = \int_{W_2} h(x)dx.$$

Therefore, if the determinant of dF is positive, then

$$\int \varphi_1^*\omega = \int \varphi_2^*\omega.$$

1. *Orientability.* Let M be a smooth manifold. Then M is *orientable* if the smooth structure can be be given by an atlas $\mathcal{A} = \{(U_\alpha, \varphi_\alpha) : \alpha \in I\}$ such that for all $\alpha, \beta \in I$ and every $x \in \varphi_\alpha^{-1}(U_\alpha \cap U_\beta)$, we have

$$\det(d(\varphi_\beta^{-1} \circ \varphi_\alpha)_x) > 0.$$

If M is orientable, then has has two *orientations*: one is given by a maximal oriented atlas $\mathcal{A} = \{(U_\alpha, \varphi_\alpha) : \alpha \in I\}$ and the other by $\mathcal{A} = \{(U_\alpha, \varphi_\alpha \circ r) : \alpha \in I\}$, where $r(x_1, x_2, \ldots, x_n) = (-x_1, x_2, \ldots, x_n)$. (Any diffeomorphism of $\mathbb{R}^n$ other than r, having negative Jacobian determinant, would do just as well.)

Exercise 7.6.1. Let M be n-dimensional. Show that if M admits a nowhere vanishing continuous n-form ω, then M is orientable. Also show the converse. (The latter requires the use of a partition of unity. This is discussed later.)

The idea of the previous exercise is that, given a maximal atlas $\mathcal{A}$, we can use ω to separate $\mathcal{A}$ into positive and negative parametrizations according to whether $\varphi^*\omega$ is given by a positive or negative multiple of $dx_1 \wedge \cdots \wedge dx_n$. Therefore, if ω defines one of two orientations of M, $-\omega$ defines the other.

CHAPTER 8

Back to probability

We establish here a general result that gives the solution to the martingale problem for a large class of hypoelliptic operators. The goal is to solve, in some appropriate sense, the differential equation

$$\dot{p} = \sum_{i=0}^{d} \dot{\omega}_i X_i(p),$$

where ω is a continuous "control curve" in $\mathbb{R}^{d+1}$ and the X_i are smooth vector fields on a manifold M.

It will be seen that a certain approximation scheme, which leads to solutions of the differential equation for smooth control, also converges as a stochastic process for almost all ω with respect to the Wienner measure on the space of continuous control curves. Furthermore, it will be seen that the probability law of the process solves the martingale problem for a certain differential operator derived from those vector fields.

1. Setting up notations

Let M be an n-dimensional manifold. Let $X_0, X_1, \ldots, X_d$ be smooth vector fields on M such that each linear combinations $X = \sum_{i=0}^{d} a_i X_i$ is a complete vector field. This is the case, for example, if the vector fields have compact support. (Recall that completeness means that the flow Φ_t^X of X is defined for all $t \in \mathbb{R}$.) We will use the notation

$$\mathcal{E}^{tX}(p) := \Phi_t^X(p).$$

Let $\mathcal{P}_0(\mathbb{R}^d)$ be the space of continuous curves $\omega : [0, \infty) \to \mathbb{R}^d$ such that $\omega(0) = 0$. (More generally, we define $\mathcal{P}_q(M)$ as the set of continuous curves in M starting at q.) It will be convenient at times to regard $\omega(t)$ as taking values in $\mathbb{R}^{d+1}$ by adding the component $\omega_0(t) = t$ to the d-dimensional vector $\omega(t) = (\omega_1(t), \ldots, \omega_d(t))$.

Given positive integers k, l and a $f : [0, \infty) \to \mathbb{R}$ we write $T_{k,l} := (k/2^l)T$ and

$$[f]_l^k(t) = f(t \wedge T_{k+1,l}) - f(t \wedge T_{k,l}).$$

We also write:

$$\begin{aligned}\Theta_{k.l}(t, \omega) &:= \sum_{i=0}^{d} [\omega_i]_l^k(t) X_i \\ &= \sum_{i=0}^{d} \left[\omega_i(t \wedge T_{k+1,l}) - \omega_i(t \wedge T_{k,l})\right] X_i.\end{aligned}$$

We now define, for each continuous ω with $\omega(0) = q$, a sequence of continuous curves on M, $p_l : t \in [0, \infty) \to M$, by the following inductive procedure. For each

l, $p_l(0) = q$, and for each $t \in [T_{k,l}, T_{k+1,l}]$, we set

$$p_l(t) = \mathcal{E}^{\Theta_{k,l}(t,\omega)}(p_l(T_{k,l})).$$

It may be convenient sometimes to write $p_l(t) = p_l(t, q, \omega)$.

a. An Example in Riemannian Geometry. Before we study the convergence properties of this approximation scheme, it will help to see what the sequence p_l corresponds to in a particular setting that will later be used to describe Brownian motion on a Riemannian manifold.

Let $(M, \langle \cdot, \cdot \rangle)$ be a Riemannian manifold of dimension d. Let $\mathcal{O}^M$ be the orthonormal frame bundle of M. Recall that to each $x \in \mathbb{R}^d$ is associated a horizontal vector field on $\mathcal{O}^M$, $Z = \mathcal{H}^x$, defined as follows. At $\sigma \in \mathcal{O}^M$, Z_σ is the unique horizontal vector in $T_\sigma \mathcal{O}^M$ that projects to $\sigma x \in T_p M$, $p = \pi(\sigma)$, under the base point map $\pi : \mathcal{O}^M \to M$. In other words, $Z_\sigma = K_\sigma \sigma x$, for each σ. We write $X_i = \mathcal{H}^{e_i}$, $i = 1, \ldots, d$, where $\{e_1, \ldots, e_d\}$ is the standard basis of $\mathbb{R}^d$.

We would like now to describe the curve obtained by projecting p_l to M. First notice the following. Let $x \in \mathbb{R}^d$ and $X = \mathcal{H}^x$ the horizontal vector field associated to x. Let $\mathcal{E}^{tX}(\sigma)$ the flow line through σ and $\gamma(t) = \pi(\mathcal{E}^{tX}(\sigma))$ the projected curve. Then $\sigma(t) := \mathcal{E}^{tX}(\sigma)$ is a horizontal left of $\gamma(t)$ to the orthonormal frame bundle. It follows that

$$\frac{\nabla}{dt}\sigma(t)y = 0$$

for every $y \in \mathbb{R}^d$. Therefore

$$\begin{aligned}\frac{\nabla}{dt}\gamma'(t) &= \frac{\nabla}{dt} d\pi_{\sigma(t)} \mathcal{H}^x(\sigma(t)) \\ &= \frac{\nabla}{dt}\sigma(t)x \\ &= 0.\end{aligned}$$

Consequently, $\gamma(t)$ is a geodesic and $\sigma(t)$ is a parallel frame along $\gamma(t)$.

The curves p_l can now be described in geometric terms. Fix an initial orthonormal frame for $T_q M$. We want to construct a curve $\bar{p}_l(t)$ on M and a frame field $\sigma(t)$ along $\bar{p}_l(t)$ such that, as a curve in $\mathcal{O}^M$, $\sigma(t) = p_l(t)$. Suppose $p_l(T_{k,l})$ has already been obtained and let $t \in [T_{k,l}, T_{k+1,l}]$. Let $\bar{p}$ be the base point of $p = p_l(T_{k,l})$. Then $p_l(t)$ is obtained by parallel translating $p_l(T_{k,l})$ to the end of the geodesic ray $s \mapsto \exp_{\bar{p}}(p[\omega_i]_l^k(s))$, $T_{k,l} \leq s \leq t$.

b. Smooth Controls. When ω is smooth, the solution to $\dot{p} = \sum_{i=0}^d \dot{\omega}_i(t) X_i$ through q is a smooth curve $p(t) = p(t, q, \omega)$ (which we know exists and is unique by standard results in ODEs) has the following characterization. For every smooth function $f : M \to \mathbb{R}$,

$$f(p(t)) = f(q) + \int_0^t \sum_{i=0}^d \dot{\omega}_i(s)(X_i f)(p(s))ds.$$

This follows immediately from the identity $f(p(t)) - f(q) = \int_0^t \frac{d}{ds} f(p(s))ds$ and the chain rule.

EXERCISE 8.1.1. The sequence $p_l(t)$ converges uniformly on compact intervals to the solution of the given initial value problem.

Here are a few indications to solving the exercise. If t lies on a sufficiently small interval, say $[0,T]$, we can assume that $p(t)$ will never leave a coordinated neighborhood of q. So it can be assumed without loss of generality that the manifold is $\mathbb{R}^n$ and $q=0$. First note that if $p_l(t)$ is a sequence of C^0 curves such that $p_l(0)=0$ and for every smooth $f:\mathbb{R}^n\to\mathbb{R}$

$$\left|f(p_l(t)) - f(0) - \int_0^t \sum_{i=0}^d \dot{\omega}_i(s)(X_i f)(p_l(s))ds\right| \to 0$$

as $n\to\infty$, for all $t\in[0,T]$, then, for every $\epsilon>0$ there exists N sufficiently big such that

$$|f(p_{l+1}(t)) - f(p_l(t))| \le \sum_{i=0}^d \int_0^t |\dot{\omega}_i(s)|\,|(X_i f)(p_{l+1}(s)) - (X_i f)(p_l(s))|\,ds + \epsilon$$

for all $l\ge N$. It follows that there exists a positive constant K (dependent on T and on the norms of the X_i on a compact neighborhood of 0) such that

$$|p_{l+1}(t) - p_l(t)| \le K\int_0^t |p_{l+1}(s) - p_l(s)|ds + \epsilon_l$$

where $\epsilon_l\to 0$ as $l\to\infty$.

By Gronwall's inequality,

$$|p_{l+1}(t) - p_l(t)| \le \epsilon_l e^{Kt}.$$

Therefore, as long as t is in a bounded interval, the above goes to 0 as $l\to\infty$. Also note that if the ϵ_l goes to 0 exponentially in l, then the p_l forms a Cauchy sequence. Therefore, if ϵ_l goes to zero fast enough the limit, $p(t)$, of the p_l would exist, and it is easy to show that $p(0)$ would be a solution to the initial value problem.

Therefore, in order to establish the proposition it is sufficient to check that for all smooth f with Lipschitz constant K_f, there exists a number $a>1$ (that depends on T and ω, but not in f or l) such that

$$\left|f(p_l(t)) - f(0) - \int_0^t \sum_{i=0}^d \dot{\omega}_i(s)(X_i f)(p_l(s))ds\right| \le K_f a^{-l}.$$

Define

$$\mathcal{D}^f_{k,l}(p) := f(\mathcal{E}^{\Theta_{k,l}(T_{k+1,l},\omega)}(p)) - f(p) - \int_{T_{k,l}}^{T_{k+1,l}} \sum_{i=0}^d \dot{\omega}_i(s)(X_i f)(\mathcal{E}^{\Theta_{k,l}(s,\omega)}(p))ds.$$

Since there are 2^l intervals of the form $[T_{k,l},T_{k+1,l}]$, if we show that

$$|\mathcal{D}^f_{k,l}(p)| \le \text{const.}b^{-l}$$

for some $b>2$, then we would get the previous inequality for an $a>1$.

Recall now Taylor's formula:

$$f(\mathcal{E}^Z(p)) - f(q) = \sum_{i=1}^k \frac{1}{i!}(Z^i f)(p) + \int_0^1 \frac{(1-s)^k}{k!}(Z^{k+1}f)(\Phi^Z_s(p))ds.$$

An application of this formula up to errors term of order 2 (in the norm of $Z=\Theta_{k,l}(t,p)$) gives

$$|\mathcal{D}^f_{k,l}(p)| \le K'(T/2^l)^2,$$

so that the needed estimate holds for $b=4$.

c. The Stochastic Case; Main Theorems. We want to show the following result. The notations are the same as in the previous section, with the following specialization: we take $\omega_0(t) = t$ and regard $\omega(t) = (\omega_1(t), \ldots, \omega_d(t))$ as an element of $\mathcal{P}_0(\mathbb{R}^d)$. It is recalled that μ denotes the Wiener measure on $\mathcal{P}_0(\mathbb{R}^d)$.

THEOREM 8.1.2 (Approximation Scheme converges). *For each $q \in M$, there exists a map*

$$p : \mathcal{P}_0(\mathbb{R}^d) \to \mathcal{P}_q(M)$$

such that, for each $T \geq 0$, $p_l(t, \omega) \to p(t, \omega) := p(\omega)(t)$ holds uniformly on $t \in [0, T]$, for μ a.e. ω. In other words, for each smooth $f : M \to \mathbb{R}$,

$$\lim_{l \to \infty} \sup_{t \in [0,T]} |f(p_l(t, q, \omega)) - f(p(t, q, \omega))| = 0$$

for almost every ω with respect to the Wiener measure.

COROLLARY 8.1.3. *The map p is $\bar{\mathcal{B}}_T$-measurable, for the completion of $\mathcal{B}_T$ with respect to the Wiener measure, where $\mathcal{B}_T$ is the σ-algebra generated by the $\pi_s : \mathcal{P}_0(\mathbb{R}^d) \to \mathbb{R}^d$, $\pi_s(\omega) = \omega(s)$, $0 \leq s \leq T$.*

Let $\mathcal{L}$ denote the second order operator:

$$\mathcal{L}f := X_0 f + \frac{1}{2} \sum_{i=1}^{d} X_i^2 f$$

where f is a smooth function on M.

THEOREM 8.1.4 (Law of p solves martingale problem). *For each $T > 0$ and each bounded smooth function $f : [0, T] \times M \to \mathbb{R}$, the map*

$$\omega \mapsto M_t^f(\omega) := f(t \wedge T, p(t \wedge T, \omega)) - \int_0^{t \wedge T} \left(\frac{\partial f}{\partial s} + \mathcal{L}f \right) (s, p(s, \omega)) ds$$

is a μ-martingale with respect to $\bar{\mathcal{B}}_t$, $t \geq 0$. In particular, the measure $\nu = p_* \mu$ on $\mathcal{P}_q(M)$ solves the martingale problem for $\mathcal{L}$, starting at q. Moreover, this measure is the unique solution to the martingale problem for $\mathcal{L}$ starting at q.

We now apply the previous two theorems to obtain a description of Brownian motion. on a Riemannian manifold $(M, \langle \cdot, \cdot \rangle)$. Let $Z_1, \ldots, Z_d$ be, as before, the canonical horizontal vector fields on $\mathcal{O}^M$, where d is the dimension of M. Let p be the limit of the family p_l, as in Theorem 8.1.2, and define the process $\bar{p} := \pi \circ p$, where $\pi : \mathcal{O}^M \to M$ is the natural projection.

In order to give a description of $\bar{p}_l$ in more Riemannian terms, let W_k be the discrete time process defined by $W_k(\omega) := p_l(T_{k,l}, \sigma_0, \omega)$, $k = 1, \ldots, 2^l$. We set $W_0(\omega) = \sigma_0$, where σ_0 is a chosen orthonormal frame above q. We are interested in the process $\overline{W}_k = \pi \circ W_k$. We also set: $\beta_t(\omega) := \omega(t)$ and $\delta_k := \beta_{T_{k,l}} - \beta_{T_{k-1,l}}$. Since $\omega(0) = 0$, we have $\delta_1(\omega) = \omega(2^{-l})$.

Using the description of $\bar{p}_l$ given earlier we see that $\overline{W}_1 = \exp_q \circ \sigma_0 \circ \delta_1$. Therefore, the law of $\overline{W}_1$ is the probability measure on M given by $(\overline{W}_1)_* \mu = (\exp_q)_* (\sigma_0)_* (\delta_1)_* \mu$. Now it is recalled that δ_1 is a Gaussian random variable with mean 0 and variance $1/2^l$. Since σ_0 is an isometry between $\mathbb{R}^d$ and $T_q M$, then $\sigma_0 \circ \delta_1$ is a Gaussian random variable with values in $T_q M$, having mean 0 and variance $1/2^l$. Moreover, the standard Gaussian distribution on $\mathbb{R}^d$ is invariant under the orthogonal group $O(d)$, so the law of $\sigma_0 \circ \delta_1$ actually does not depend on the choice of the initial frame, σ_0. Therefore, the random variable $\overline{W}_1$ is simply the exponential of a normally distributed random vector in T_qM, with mean 0 and variance $1/2^l$.

Having obtained $q_{k-1} = \overline{W}_{k-1}(\omega)$, $\overline{W}_k$ is now the exponential of a normally distributed random vector in $T_{q_{k-1}}M$, always with mean 0 and variance $1/2^l$. Notice, in particular, that the law of the process $\bar{p}$ does not depend on the choice of initial frame on T_qM.

Finally, we claim that the law of $\bar{p}$ solves the martingale problem for the Riemannian Laplacian operator, Δ, with initial point q. This is an immediate consequence of Theorem 8.1.4 and a fact about the Laplacian that will be recalled shortly. In fact, let $f \in C^0([0,\infty) \times M; \mathbb{R})$ be a function that is smooth on M for each $t \in [0,\infty)$. Consider the function $g(t,\sigma) = f(t,\pi(\sigma))$, $\sigma \in \mathcal{O}^M$. According to Theorem 8.1.4,

$$\omega \mapsto M_t^g(\omega) := g(t \wedge T, p(t \wedge T, \omega)) - \int_0^{t\wedge T} \left(\frac{\partial g}{\partial s} + \mathcal{L}g\right)(s, p(s,\omega))ds$$

is a μ-martingale with respect to $\bar{\mathcal{B}}_t$, $t \geq 0$, where $\mathcal{L} = \sum_{i=1}^d Z_i^2$. We know, on the other hand, that $(\mathcal{L}g)(t,\sigma) = (\Delta f)(t,\pi(\sigma))$. Consequently,

$$\omega \mapsto M_t^f(\omega) = f(t \wedge T, \bar{p}(t \wedge T, \omega)) - \int_0^{t\wedge T} \left(\frac{\partial f}{\partial s} + \Delta f\right)(s, \bar{p}(s,\omega))ds$$

also is a martingale since $M_t^f(\omega) = M_t^g(\omega)$.

d. Further Properties of p. It is possible to find a limit p for the sequence p_l such that p depends smoothly on the initial point $q \in M$. The precise statement is the following.

THEOREM 8.1.5 (Smooth Dependence on Initial Point). *The same assumptions and notations of Theorem 8.1.2 are in force. There exists a measurable map*

$$\omega \in \mathcal{P}_0(\mathbb{R}^d) \mapsto p(\cdot,\cdot,\omega) \in C^0([0,\infty) \times M; M)$$

such that $q \mapsto p(t,q,\omega)$ is a C^∞ function, and for every smooth compactly supported function $f : M \to \mathbb{R}$,

$$\lim_{l\to\infty} \sup_{t\in[0,T]} \sup_{q\in M} |\partial_q^\beta f(p_l(t,q,\omega)) - \partial_q^\beta f(p(t,q,\omega))| = 0$$

both μ-almost surely and in $L^2(\mathcal{P}_0(\mathbb{R}^d,\mu)$. Moreover,

$$(t,q) \in [0,\infty) \times M \mapsto \int_{\mathcal{P}_0(\mathbb{R}^d)} f(p(t,q,\omega))d\mu(\omega) \in \mathbb{R}$$

is a continuous function that depends smoothly on q.

Notice that if f is a twice differentiable bounded function on M that does not depend explicitly on t, then Theorem 8.1.4 implies:

$$E[f(p(T,q,\cdot))] = f(q) + E\left[\int_0^T (\mathcal{L}f)(p(s,q,\cdot))ds\right]. \tag{8.1}$$

There is a very useful generalization of Equation 8.1 known as Dynkin's formula. We first need the definition of *stopping times*. A function $\zeta : \mathcal{P}_0(\mathbb{R}^d) \to [0,\infty]$ is said to be a stopping time with respect to the filtration $\bar{\mathcal{B}} := \{\bar{\mathcal{B}}_t : t > 0\}$ if

$$\{\omega \in \mathcal{P}_0(\mathbb{R}^d) : \zeta(\omega) \leq t\} \in \bar{\mathcal{B}}_t$$

for each $t \in [0,\infty)$.

Theorem 8.1.6 (Dynkin's Formula). *Let f be a bounded twice differentiable function on M and ζ a stopping time with respect to $\bar{\mathcal{B}}$ such that $E[\zeta] < \infty$, where E is expectation with respect to the law $\nu = p_*\mu$ (with respect to which almost all sample paths start at q). Then*

$$E[f(p(\zeta, q, \cdot))] = f(q) + E\left[\int_0^\zeta (\mathcal{L}f)(p(s, q, \cdot))ds\right].$$

Notice the following immediate application of Dynkin's formula. Suppose that D is an open connected subset of a manifold M and $\mathcal{L}$ is as before. For each $q \in D$, let $\zeta = \zeta^q$ denote the first exit time from D. (This is, for each ω, the infimum of t such that $p(t, q, \omega)$ lies in the complement of D.) The map

$$\omega \in \mathcal{P}_0(\mathbb{R}^d) \mapsto p(\zeta(\omega), q, \omega) \in \partial D$$

pushes the Wiener measure μ to a probability measure η^q on ∂D. Suppose now that $f : D \to \mathbb{R}$ is $\mathcal{L}$-harmonic, that is. f is twice differentiable and $\mathcal{L}f = 0$ on D. Let ϕ denote the boundary values of f. Then, since $f(q) = E[f(p(\zeta, q, \cdot))]$, we can write

$$f(q) = \int_{\partial D} \phi(p) d\eta^q(p).$$

In other words, an $\mathcal{L}$-harmonic function on D is determined by its values at the boundary ∂D. The measure η^q is called the *harmonic measure* of the process $p(t, q, \cdot)$ on ∂D. This idea can be used to obtain existence and uniqueness of solution of for the Dirichlet problem on D (under appropriate hypothesis).

The notation $(\delta_s\omega)(t) := \omega(s + t) - \omega(s)$ will be used below.

Theorem 8.1.7 (Renewal at Stopping Times). *Let $\zeta : \mathcal{P}_0(\mathbb{R}^d) \to [0, \infty]$ be a stopping time and p as in Theorem 8.1.2. Then*

$$p(t + \zeta(\omega), q, \omega) = p(t, p(\zeta(\omega), q, \omega), \delta_{\zeta(\omega)}\omega)$$

for $t \in [0, \infty)$ and μ-a. e. ω.

In the next corollary we use E^q to denote expectation with respect to the measure for which almost all sample paths start at q.

Corollary 8.1.8 (The Strong Markov Property). *Let f be a bounded function on M and ζ a stopping time with respect to $\bar{\mathcal{B}}$. Then*

$$E^q[f(p(t + \zeta, q, \cdot))|\bar{\mathcal{B}}_\zeta] = E^{p(\zeta,q,\cdot)}[f(p(t, p(\zeta, q, \cdot), \cdot))].$$

Let $f : M \to \mathbb{R}$ be a bounded measurable function and define

$$[P_t^{\mathcal{L}} f](q) := E^\mu[f(p(t, q, \cdot))].$$

The "Renewal at Stopping Times" theorem, applied to a constant ζ, says that the family of operators $P_t^{\mathcal{L}}$, $t \in [0, \infty)$, forms a 1-parameter semigroup:

$$P_{t+s}^{\mathcal{L}} = P_t^{\mathcal{L}} \circ P_s^{\mathcal{L}}$$

Moreover, if f is a twice differentiable bounded function on M, then the function $u(t, q) := [P_t^{\mathcal{L}} f](q)$ satisfies

$$\frac{\partial u}{\partial t}(t, q) = \mathcal{L}u(t, q)$$

on $[0, \infty) \times M$, with initial condition $u(0, \cdot) = f$.

2. Proof of the First Main Theorem

Here are, once again, the main notations and definitions (including a few new ones): $X_0, X_1, \ldots, X_d$ are smooth, compactly supported vector fields on the smooth n-dimensional manifold M. We fix throughout $T > 0$ and $q \in M$. Recall that $T_{k,l} = kT/2^l$.

It will be convenient to introduce the *increment process* $[\cdot]_l^k(t) : \mathcal{P}_0(\mathbb{R}^d) \to \mathbb{R}^d$ defined by

$$\omega \mapsto [\cdot]_l^k(t)(\omega) := \ [\omega]_l^k(t) := \omega(t \wedge T_{k+1,l}) - \omega(t \wedge T_{k,l}).$$

With this notation, the previously defined vector field $\Theta_{k,l}(t, \omega)$ takes the form

$$\Theta_{k,l}(t, \omega) = \sum_{i=0}^{d} [\omega_i]_l^k(t) X_i.$$

We are adding, here, the extra component, $\omega_0(t) = t$.

The process p_l is, once again, defined inductively by: $p_l(0) = q$ and

$$p_l(t) = \mathcal{E}^{\Theta_{k,l}(t)}(p_l(T_{k,l}))),$$

for $T_{k,l} \leq t \leq T_{k+1,l}$,

We fix now a compactly supported smooth function $f : M \to \mathbb{R}$. The proof will revolve around finding estimates for the quantity

$$\delta_l^f(t) := f(p_{l+1}(t)) - f(p_l(t))$$

so as to show that $f(p_l(t))$ converges. (Equivalently, we could estimate the quantity

$$\delta_l(t) := \sup\{\delta_l^f(t) : f \text{compactly supported, 1-Lipschitz function on } M\},$$

where the Lipschitz condition is with respect to some (arbitrary) choice of Riemannian metric and associate distance function d.)

Exercise 8.2.1. If M is given a Riemannian metric and $p_1, p_2 \in M$, show that

$$\sup\{|f(p_1) - f(p_2)| : f \text{compactly supported, 1-Lipschitz function on } M\} = d(p_1, p_2),$$

where $d(p_1, p_2)$ is the Riemannian distance function. (Clearly, the notion of 1-Lipschitz function depends on the choice of metric.)

More precisely, we are interested in showing that the quantity

$$E\left[\sup_{t\in[0,T]} |f(p_{l+1}(t)) - f(p_l(t))|^2\right] := \int_{\mathcal{P}_0(\mathbb{R}^d)} \left(\sup_{t\in[0,T]} |f(p_{l+1}(t)) - f(p_l(t))|^2\right) d\mu$$

goes to zero fast enough (exponentially) with l. The conclusion of Theorem 8.1.2 will then follow by general facts of measure theory.

The observation contained in the following exercise will be important. Let Z_1, Z_2 be vector fields on M which are linear combinations of the X_i, the coefficients being constants. (More precisely, the coefficients do not depend of the point in M, although they may depend on time.) Then, an immediate application of the general (Taylor) formula

$$f(\mathcal{E}^Z(p)) = f(p) + (Zf)(p) + \cdots + \frac{1}{k!}(Z^k f)(p) + \frac{1}{k!}\int_0^1 (1-t)^k (Z^{k+1} f)(\mathcal{E}^{tZ}(p))dt,$$

implies the next exercise. (Notice that if $a_Z = (a_0, \ldots, a_d)$ is the vector of coefficients of $Z = \sum_{i=0}^d a_i X_i$, then the remainder term is a homogeneous polynomial in a_Z of degree $k+1$.)

EXERCISE 8.2.2. Let Z_1, Z_2 be as just defined. Then, for each $p \in M$, show that

$$f(\mathcal{E}^{Z_1}(\mathcal{E}^{Z_2}(p))) = f(\mathcal{E}^{Z_1+Z_2}(p)) + \frac{1}{2}[Z_2, Z_1]_p f + \text{third order terms in coefficients of } Z_i.$$

Also note the fact stated in the next exercise, which is an immediate consequence of definitions.

EXERCISE 8.2.3. For $T_{k,l} \le t \le T_{k+1,l}$, show that

$$f(p_{l+1}(t)) = f(\mathcal{E}^{\Theta_{2k+1,l+1}(t)}(\mathcal{E}^{\Theta_{2k,l+1}(t)}(p_{l+1}(T_{k,l})))).$$

EXERCISE 8.2.4. Check that $X(t) + Y(t) = Z(t)$ for all $t \ge 0$, where

$$\begin{aligned} X(t) &:= \Theta_{2k+1,l+1}(t) \\ Y(t) &:= \Theta_{2k,l+1}(t) \\ Z(t) &:= \Theta_{k,l}(t). \end{aligned}$$

Let $X(t), Y(t), Z(t)$ be as in the exercise. Also set

$$p_1 := p_{l+1}(T_{k,l}), \quad p_2 := p_l(T_{k,l})$$

and let $a_{Z(t)}$ (respectivelly, $a_{X(t)}$ and $a_{Y(t)}$) denote the vector in $\mathbb{R}^{d+1}$ whose components are the coefficients of $Z(t)$ (respectivelly, $X(t)$ and $Y(t)$) with respect to the vector fields X_i. (Incidentally, we are not assuming that the X_i are linearly independent.) Then, for $t \in [T_{k.l}, T_{k+1,l}]$, we have

$$\begin{aligned} \delta_l^f(t) - \delta_l^f(T_{k,l}) &= f(p_{l+1}(t)) - f(p_1) - (f(p_l(t)) - f(p_2)) \\ &= f(\mathcal{E}^{X(t)}(\mathcal{E}^{Y(t)}(p_1))) - f(p_1) - \Big(f(\mathcal{E}^{X(t)+Y(t)}(p_2)) - f(p_2)\Big) \\ &= f(\mathcal{E}^{X(t)+Y(t)}(p_1)) + \frac{1}{2}[Y(t), X(t)]_{p_1} f + r_2(a_{X(t)}, a_{Y(t)})_{p_1} \\ &\qquad - \Big(f(\mathcal{E}^{X(t)+Y(t)}(p_2)) - f(p_2)\Big) \\ &= (Z(t)f)(p_1) - (Z(t)f)(p_2) + \frac{1}{2}[X(t), Y(t)]_{p_1} f \\ &\qquad + r_2(a_{X(t)}, a_{Y(t)})_{p_1} + r_1(a_{X(t)})_{p_1} - r_2(a_{Y(t)})_{p_2} \\ & A_l^k(t) + B_l^k(t) + C_l^k(t) + D_l^k(t), \end{aligned}$$

where (leaving out explicit reference to t)

$$\begin{aligned} A_l^k &= \sum_{i=1}^d \left((Z^i f)(p_1) - (Z^i f)(p_2)\right); \\ B_l^k &= ((Z^0 f)(p_1) - (Z^0 f)(p_2)) + r_1(a_Z)_{p_1} - r_1(a_Z)_{p_1}; \\ C_l^k &= \frac{1}{2} \sum_{i,j=1}^d [Y^i, X^j]_{p_1} f; \\ D_l^k &= \frac{1}{2} \sum_{i=1}^d \left([Y^0, X^i]_{p_1} + [Y^i, X^0]_{p_1}\right) f + r_2(a_X, a_Y)_{p_1} + \frac{1}{2}[Y^0, X^0]_{p_1} f. \end{aligned}$$

(Here, X^i, Y^i, Z^i denote $(a_X)_iX_i$, $(a_Y)_iX_i$, $(a_Z)_iX_i$, respectively.)

Note that $\delta_l^f(0) = 0$ and that, for $t \in [T_{k.l}, T_{k+1,l}]$, we have

$$\begin{aligned}\delta_l^f(t) &= \delta_l^f(t) - \delta_l^f(T_{k,l}) + \delta_l^f(T_{k,l}) \\ &= \delta_l^f(t) - \delta_l^f(T_{k,l}) + \sum_{m=0}^{k-1} \left(\delta_l^f(T_{m+1,l}) - \delta_l^f(T_{m,l})\right) \\ &= A_l(t) + B_l(t) + C_l(t) + D_l(t),\end{aligned}$$

where

$$A_l(t) = \sum_{k=0}^{\infty} A_l^k(t),\ B_l(t) = \sum_{k=0}^{\infty} B_l^k(t),\ C_l(t) = \sum_{k=0}^{\infty} C_l^k(t),\ D_l(t) = \sum_{k=0}^{\infty} D_l^k(t).$$

We need now to muster the courage to evaluate:

$$E\left[\sup_{t\in[0,T]} |\delta_l^f(t)|^2\right] := E\left[\sup_{t\in[0,T]} |A_l(t) + B_l(t) + C_l(t) + D_l(t)|^2\right].$$

Since $(A+B+C+D)^2 \le 4(A^2+B^2+C^2+D^2)$ for any real numbers A, B, C, D, we are left with the problem of estimating the quantities:

$$E\left[\sup_{t\in[0,T]} |A_l(t)|^2\right], E\left[\sup_{t\in[0,T]} |B_l(t)|^2\right], E\left[\sup_{t\in[0,T]} |C_l(t)|^2\right], E\left[\sup_{t\in[0,T]} |D_l(t)|^2\right].$$

We denote these expectations by $\bar{A}_l, \bar{B}_l, \bar{C}_l, \bar{D}_l$, respectively.

In a first step, we would like to bound each of the above expression using (sums over k of) the expectations of $\delta_l(T_{k,l})$. We will then apply a form of Gronwall's inequality to arrive at the desired estimate for $E\left[\sup_{t\in[0,T]} |\delta_l^f(t)|^2\right]$.

A key ingredient in the calculation will be the $[\omega_i]_l^m(t)$ are martingales and that $[\omega_i]_l^m(t)$ and $[\omega_i]_l^{m'}(t)$ are independent when $m \neq m'$. Before starting with the details we state the following important inequality. A proof of it will be given later.

Theorem 8.2.5 (Doob's Inequality). *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{F}_t$, $t \ge 0$, a filtration of sub σ-algebras of $\mathcal{F}$. Let $M(t)$ be a (right-)continuous $\mathbb{R}$-valued martingale with respect to $\mathcal{F}_t$. Then, for each $p > 1$ and ≥ 0,*

$$E\left[\sup_{t\in[0,T]} |M(t)|^p\right]^{\frac{1}{p}} \le \frac{p}{p-1} E\left[|M(T)|^p\right]^{\frac{1}{p}}.$$

a. Bounding $\bar{A}_l$. We first observe that

$$A_l^k(t) = \sum_{i=1}^{d} [\omega_i]_l^k(t) \left((X_if)(p_{l+1}(T_{k,l})) - (X_if)(p_l(T_{k,l}))\right)$$

is a martingale for the filtration $\{\mathcal{B}_t : t \ge 0\}$ and the Wiener measure μ on $\mathcal{P}_0(\mathbb{R}^d)$. This is immediate since $[\omega_i]_l^k(t)$ is a martingale. In particular, the sum $A_l(t)$ is also a martingale.

By Doob's Inequality, we have

$$E\left[\sup_{t\in[0,T]} |A_l(t)|^2\right] \le 4E\left[|A_l(T)|^2\right],$$

so that

$$\bar{A}_l \leq 4E\left[\left(\sum_{k=0}^{[2^lT]} A_l^k(T)\right)^2\right] = 4\sum_{k,k'=0}^{[2^lT]} E\left[A_l^k(T)A_l^{k'}(T)\right].$$

EXERCISE 8.2.6. Show that

$$E\left[[\omega_i]_l^k(t)[\omega_j]_l^{k'}(t)\right] = \begin{cases}[\omega_0]_l^k(T) & \text{if } i=j \text{ and } k=k'\\ 0 & \text{otherwise.}\end{cases}$$

Using the previous exercise, we obtain:

$$\bar{A}_l \leq C_A \sum_{k=0}^{[2^lT]} [\omega_0]_l^k(T)E[\delta_l(T_{k,l})^2].$$

b. Bounding $\bar{B}_l$. We have that $B_l^k = \beta_l^k + \bar{\beta}_l^k$ where

$$\beta_l^k = [\omega_0]_l^k(t)\left((X_0f)(p_{l+1}(T_{k,l})) - (X_0f)(p_l(T_{k,l}))\right)$$

$$\bar{\beta}_l^k = \sum_{i,j=0}^{d} [\omega_i]_l^k(t)[\omega_j]_l^k(t)(r_{(i,j)}(Z(t))_{p_{l+1}(T_{k,l})} - r_{(i,j)}(Z(t))_{p_l(T_{k,l})}).$$

The error term $r_1(a_Z)_p$ is written

$$r_{(i,j)}(Z(t))_p := \int_0^1 (1-s)(X_iX_jf)(\mathcal{E}^{sZ}(p))ds.$$

(A similar expansion can be written for $r_2(a_X,a_Y)_{p_1}$. It would give a homogeneous third degree polynomial in the various $[\omega_i]_l^m(t)$.)

It follows that

$$|B_l^k(t)| \leq C_B^{(1)}\left([\omega_0]_l^k(t) + |[(\omega_0,\ldots,\omega_d)]_l^k(t)|^2\right)\delta_l(T_{k,l}).$$

We now use the fact that $[\omega_0]_l^k(t) + |[(\omega_0,\ldots,\omega_d)]_l^k(t)|^2$ is independent of $\delta_l(T_{k,l})$ to obtain (Exercise 8.2.6 is also used):

$$E\left[\sup_{0\leq t\leq T}|B_l^k(t)|^2\right] \leq C_B^{(2)}\left([\omega_0]_l^k(T)\right)^2 E[(\delta_l(T_{k,l}))^2].$$

Now, observe the following. By the Schwarz's inequality we have

$$\left(\sum_{k=0}^{[2^lT]} B_l^k(t)\right)^2 \leq (1+[2^lT])\sum_{k=0}^{[2^lT]}(B_l^k(t))^2.$$

Taking the supremum, and then expectations, we get

$$\begin{aligned}E\left[\sum_{0\leq t\leq T}(B_l(t))^2\right] &\leq (1+[2^lT])\sum_{k=0}^{[2^lT]} E\left[\sup_{0\leq t\leq T}(B_l^k(t))^2\right]\\ &\leq (1+[2^lT])\sum_{k=0}^{[2^lT]} C_B^{(2)}\left([\omega_0]_l^k(T)\right)^2 E[(\delta_l(T_{k,l}))^2]\\ &\leq C_B\sum_{k=0}^{[2^lT]}[\omega_0]_l^k(T)E[(\delta_l(T_{k,l}))^2],\end{aligned}$$

where $C_B = C_B^{(2)}$.

c. Bounding $\bar{C}_l$. Observe that

$$C_l^k(t) = \frac{1}{2} \sum_{i,j=1}^{d} [\omega_i]_{l+1}^{2k}(t)[\omega_j]_{l+1}^{2k+1}(t)[X_i, X_j]_{p_{l+1}(T_{k,l})} f$$

is also a martingale, since $[\omega_i]_{l+1}^{2k}(t)$ and $[\omega_j]_{l+1}^{2k+1}(t)$ are independent. Using again Exercise 8.2.6 it can be shown that

$$\bar{C}_l \leq C_C \sum_{k=0}^{[2^l T]} \left([\omega_0]_l^k(T)\right)^2.$$

d. Bounding $\bar{D}_l$. A similar argument can be applied to

$$D_l^k(t) = \frac{1}{2} \sum_{i=1}^{d} \left([\omega_0]_{l+1}^{2k}(t)[\omega_i]_{l+1}^{2k+1}(t) - [\omega_0]_{l+1}^{2k+1}(t)[\omega_i]_{l+1}^{2k}(t)\right) [X_0, X_i]_{p_1} f + r_2(a_X, a_Y).$$

The outcome is that

$$E\left[\sup_{0 \leq t \leq T} |D_l|^2\right] \leq C_D \sum_{k=0}^{[2^l T]} \left([\omega_0]_l^k(T)\right)^2.$$

(Here, $C_D = C_C T$.)

e. Putting it all together. Adding the four terms and taking the supremum over 1-Lipschtz funtions yields

$$E\left[\sup_{0 \leq t \leq T} |\delta_l(t)|^2\right] \leq \frac{C}{2^l} \left(T + \sum_{k=0}^{[2^l T]} E\left[\sup_{0 \leq t \leq T} |\delta_l(t)|^2\right]\right).$$

EXERCISE 8.2.7. Suppose that a finite sequence $u_1, \ldots, u_N$ satisfies:

$$u_l \leq a \sum_{n=0}^{l} u_n + b$$

for constants a, b and all $n = 1, \ldots, N$. Show that $u_N \leq (au_0 + b)e^{aN}$.

The exercise implies that for some constant C

$$E\left[\sup_{0 \leq t \leq T} |\delta_l(t)|^2\right] \leq \frac{C}{2^l}.$$

It is now easy to conclude that for each N there exists a constant $K > 0$ such that

$$E\left[\sup_{n \geq N} \sup_{0 \leq t \leq T} |p_l(t) - p_N(t)|^2\right] \leq \frac{K}{2^N}.$$

But this is what was needed to show that the approximation scheme converges.

f. Proof that limit of p_l solves Martingale property. We now turn to the proof of the second main theorem. Let $f : [0,\infty) \times M \to \mathbb{R}$ be a smooth compactly supported function. Let $t_1 = T_{k_1,N} < t_2 = T_{k_2,N} \le T$ and fix $L \ge N$. Write $g_L(t) := f(t, p_L(t))$. Then

$$g(t_2) - g(t_1) = \sum_{k=k_1 2^{L-N}}^{k_2 2^{L-N}-1} \left(g(T_{k+1,L} - g(T_{k,L})\right).$$

The Taylor approximation of each summand is (f_1 denotes the partial derivative of f in the first argument):

$$\begin{aligned} g(t) - g(T_{k,L}) = \int_{T_{k,L}}^{t} f_1(s, p_L(T_{k,L}))ds + \sum_{i=0}^{d} [\omega_i]_L^k(t)(X_i f)(t, p_L(T_{k,L})) + \\ \frac{1}{2}\sum_{i,j=1}^{d} [\omega_i]_L^k(t)[\omega_j]_L^k(t)(X_j X_i f)(t, p_L(T_{k,L})) + \\ \text{error term in } \left\{[\omega_0]_L^k(t)[\omega_i]_L^k(t), [\omega_{i_1}]_L^k(t)[\omega_{i_2}]_L^k(t)[\omega_{i_3}]_L^k(t)\right\}. \end{aligned}$$

Taking conditional expectation with respect to the σ-algebra $\overline{\mathcal{B}}_{T_{k,L}}$ and using Exercise 8.2.6, we obtain for $T_{k,L} \le t \le T_{k+1,L}$:

$$\begin{aligned} E[g(t) - g(T_{k,L})|\overline{\mathcal{B}}_{T_{k,L}}] = \int_{T_{k,L}}^{T_{k+1,L}} f_1(s, p_L(T_{k,L}))ds + \\ [\omega_0]_L^k(t)\left(\mathcal{L}f\right)(t, p_L(T_{k,L})) + R_L, \end{aligned}$$

where $\mathcal{L} = X_0 + \frac{1}{2}\sum_{i=1}^d X_i^2$ and R is a term dominated by $C2^{-3L/2}$. It follows that

$$\begin{aligned} E\left[g(t_2) - g(t_1)|\overline{\mathcal{B}}_{t_1}\right] = E\left[\sum_{k=k_1 2^{L-N}}^{k_2 2^{L-N}-1} \int_{T_{k,L}}^{T_{k+1,L}} f_1(s, p_L(T_{k,L}))ds \middle| \overline{\mathcal{B}}_{t_1}\right] + \\ \sum_{k=k_1 2^{L-N}}^{k_2 2^{L-N}-1} E\left[\frac{1}{2^L}\left(\mathcal{L}f\right)(t, p_L(T_{k,L})) + R_L \middle| \overline{\mathcal{B}}_{t_1}\right]. \end{aligned}$$

Passing to the limit $L \to \infty$, we obtain

$$E\left[f(t_2, p(t_2)) - f(t_1, p(t_1))|\overline{\mathcal{B}}_{t_1}\right] = E\left[\int_{t_1}^{t_2}\left(\frac{\partial f}{\partial s} + \mathcal{L}f\right)(s, p(s))ds \middle| \overline{\mathcal{B}}_{t_1}\right].$$

But this is precisely what we wanted to show.

3. Mean Exit Time from a Disc

Let M be a smooth manifold with Riemannian metric $\langle\cdot,\cdot\rangle$. The space of continuous curves $\gamma : [0,\infty) \to M$ (no initial point specified) will be written $\mathcal{P}(M)$. In the present section we would like to study the function $\zeta_r : \mathcal{P}(M) \to [0,\infty]$ defined by

$$\zeta_r(\gamma) := \inf\{t \ge 0 : \mathrm{dist}^M(p(t), p(0)) \ge r\},$$

where r is a positive number. In particular, we will be interested to find the expected value of ζ_r with respect to the Wiener measure. That is, we would like to find the mean time it takes for Brownian motion to travel a distance r from its point of departure.

a. The Euclidian case first. Before considering the general case, suppose that $M = \mathbb{R}^n$ with the ordinary euclidian metric. Therefore Δ will be the ordinary Laplacian. Since the group of isometries is transitive on $\mathbb{R}^n$, we need only consider the expected value of ζ_r on $\mathcal{R}_0(\mathbb{R}^n)$ (with the Wiener measure).

We consider the slightly more general setting. Let $K = B(r, 0)$ be the ball of radius r centered at the origin, and fix a point q in the interior of K. We want to know the expected exit time, τ_K^q, from K of Brownian motion starting at q.

Choose an interger N and apply Dynkin's formula to the function $h(x) = |x|^2$ and stopping time $\tau_N := N \wedge \tau_K^q$:

$$E[h(p(\tau_N, q, \cdot))] = h(q) + E\left[\int_0^{\tau_N} (\frac{1}{2}\Delta h)(p(s, q, \cdot))ds\right].$$

But the Laplacian of $|x|^2$ is $2n$, so the identity reduces to

$$E[|p(\tau_N, q, \cdot)|^2] = |q|^2 + nE[\tau_N].$$

In particular, (since $h(p(\tau_N, q, \cdot)) \leq r^2$) $E[\tau_N] \leq r^2 - |q|^2$ is uniformly bounded on N. Passing to the limit as $N \to \infty$ gives:

$$E[\tau_K^q] = \frac{1}{n}(r^2 - |q|^2).$$

It follows that ζ_r has expected value r^2/n.

b. Remark on Recurrence and Transience. Suppose now that the inital point q is outside K, that is, $|q| > r$. We would like to know the probability that Brownian motion starting at q will ever reach K.

Let ζ_k be the first exit time from the annulus

$$A_k = \{x \in \mathbb{R}^n : r < |x| < 2^k r\}.$$

Let $f = f_{n,k}$ be a twice differentiable function with compact support such that, for $r \leq |x| \leq 2^k r$,

$$f(x) = h(|x|) = \begin{cases} -\log|x|, & \text{if } n = 2 \\ |x|^{2-n} & \text{if } n > 2. \end{cases}$$

Then f is harmonic in A_k and we have, by Dynkin's formula (for the Wiener measure on $\mathcal{P}_q(\mathbb{R}^n)$):

$$E[f(B_{\zeta_k})] = f(q).$$

Let p_k denote the probability of the event $|B_{\zeta_k}| = r$, and $q_k = 1 - p_k$ the probability of the event $|B_{\zeta_k}| = 2^k r$. Since $E[f(B_{\zeta_k})] = h(r)p_k + h(2^k r)q_k$, we obtain: $h(|q|) = h(r)p_k + h(2^k r)q_k$.

When $n = 2$, the previous formula reduces to

$$-\ln|q| = -p_k \log r - q_k \ln(2^k r).$$

It follows that $q_k \to 0$ as $k \to \infty$, so that $p_k \to 1$. In other words, the probability that Brownian motion that starts at q (outside K) will hit K before going out to infinity is 1. In this case (of Brownian motion on $\mathbb{R}^2$), the process is said to be *recurrent*.

Suppose now that the dimension, n, is greater than 2. We have

$$p_k r^{2-n} + q_k (2^k r)^{2-n} = |q|^{2-n}.$$

It follows that

$$\lim_{k\to\infty} p_k = \left(\frac{|q|}{r}\right)^{2-n} < 1.$$

Therefore, with probability greater than 0, Brownian motion will escape to infinity without hitting K. The process is said to be *transient* in this case.

EXERCISE 8.3.1. Suppose that $|q| > r$ and $n \geq 2$. Show that the expected time it takes Brownian motion to hit $B(0, r)$ having started at q is ∞.

c. On to Riemannian Manifolds. We would like to estimate the expected value of ζ_r, for small r, on a general Riemannian manifold. We will find that

$$E[\zeta_r] = \frac{r^2}{n} + \frac{r^4}{6n^2(n+2)} S(q) + o(r^4),$$

where $S(q)$ is the scalar curvature at q, defined by

$$S(q) = \frac{1}{n(n-1)} \sum_{i,j} \langle R(e_i, e_j)e_j, e_i\rangle,$$

where $e_1, \ldots, e_n$ is an arbitrary orthonormal basis of T_qM.

Some work will be needed to find approximations of the Laplacian in a normal coordinate system.

Let M be an n-dimensional Riemannian manifold with Riemannian metric $\langle \cdot, \cdot \rangle$. Fix $q \in M$, e_i, $i = 1, \ldots, n$, an orthonormal basis of T_qM, and let Y_i, $i = 1, \ldots, n$ denote the coordinate vector fields associated to the normal coordinate system

$$\varphi : (t_1, \ldots, t_n) \mapsto \exp_q(t_1 e_1 + \cdots + t_n e_n).$$

For each $u = u_1 e_1 + \cdots + u_n e_n \in T_qM$ set

$$A_u := d(\exp_q)_u : T_qM \to T_pM.$$

(We are using here the natural identification of $T_u(T_qM)$ with T_qM.) Notice that, if $p = \exp_q u$, then

$$Y_i(p) = \frac{\partial \varphi}{\partial t_i}(u_1, \ldots, u_n) = A_u e_i.$$

Therefore, if $A_u^t : T_pM \to T_qM$ is the adjoint of A_u with respect to the inner products at p and q then

$$g_{ij}(p) := \langle Y_i(p), Y_j(p)\rangle_p = \langle A_u^t A_u e_i, e_j\rangle_q.$$

In particular,

$$\mathcal{J}(p) := \det A_u = \sqrt{\det(g_{ij})_p}.$$

(The determinant of A_u is positive for u near 0 since $A_0 = I$.)

1. *An interesting fact about* Δ. We show here the following fact:

PROPOSITION 8.3.2. *Let $f : M \to \mathbb{R}$ be a smooth function which is radial on a coordinate neighborhood of q. In other words, for $p = \exp_q u$, $f(p) = h(|u|)$. Denote by $\frac{\partial}{\partial r}$ the radial coordinate vector field around q, where $r(p) = |u|$. Then, at p,*

$$\Delta f = \frac{\partial^2 f}{\partial r^2} + \left(\frac{n-1}{r} + \frac{\partial}{\partial r}\log \mathcal{J}\right)\frac{\partial f}{\partial r}.$$

In particular, if $f(p) = \frac{1}{2} r(p)^2$, then

$$\Delta f = n + r\frac{\partial}{\partial r}\log \mathcal{J}.$$

Before giving the proof of the proposition, we need to discuss some general facts. Denote $R = \frac{\partial}{\partial r}$. If f is radial, we also write $Rf = f'$.

EXERCISE 8.3.3. If f is radial near q, show the following identities:

$$\begin{aligned} r\mathrm{div}R &= \mathrm{div}(rR) - 1; \\ \mathrm{grad}f &= f'R; \\ \Delta f &= f'' + (\mathrm{div}R)f'. \end{aligned}$$

Therefore, in order to establish the formula for the Laplacian, it suffices to prove that the vector field $\mathcal{R} := rR$ satisfies:

$$\mathrm{div}\mathcal{R} = n + \mathcal{R}\log\mathcal{J}.$$

EXERCISE 8.3.4. This exercise describes a few properties of the radial vector field $\mathcal{R}$.

(1) Show that

$$\mathcal{R} = \sum_{i=1}^{n} y_i Y_i,$$

where the y_i represent the coordinate functions for the normal coordinate system.

(2) Use the previous fact to show that

$$[\mathcal{R}, Y_i] = -Y_i$$

for each $i = 1, \ldots, n$.

(3) If $p(y)$ is a homogeneous polynomial of degree k in the normal coordinates $y_1, \ldots, y_n$, show that

$$\mathcal{R}p(y) = kp(y).$$

EXERCISE 8.3.5. Let $t \mapsto A(t)$ is a differentiable matrix-valued function such that $A(t)$ is a square real matrix with positive determinant for each t. Show that

$$\frac{d}{dt}(\log\det A(t)) = \mathrm{Trace}\left(A(t)^{-1}\frac{d}{dt}A(t)\right).$$

We fix a ray $t \mapsto \gamma(t) = \exp_q(tu)$ and consider a parallel orthonormal frame $e_1(t), \ldots, e_n(t)$ along $\gamma(t)$ such that $e_n(t) = \gamma'(t)$. (Therefore, $u = e_n(0)$.) Let $A(t) = (a_{ij}(t))$ be the matrix such that

$$a_{ij}(t) = \langle A_{te_n} e_i(0), e_j(t)\rangle.$$

We also write $A(t)^{-1} = (\bar{a}_{ij})$. Notice that $\mathcal{J}(\gamma(t)) = \det A(t)$. Therefore,

$$
\begin{aligned}
\frac{d}{dt}\log\mathcal{J}(\gamma(t)) &= \operatorname{Trace}(A(t)^{-1}A'(t)) \\
&= \sum_{i=1}^{n}\left\langle A_{te_n(0)}^{-1}\frac{\nabla}{dt}A_{te_n(0)}e_i(0), e_i(0)\right\rangle \\
&= \frac{1}{t}\sum_{i=1}^{n}\left\langle A_{te_n(0)}^{-1}(\nabla_{\mathcal{R}}Y_i)_{\gamma(t)}, e_i(0)\right\rangle \\
&= \frac{1}{t}\sum_{i=1}^{n}\left\langle A_{te_n(0)}^{-1}(\nabla_{Y_i}\mathcal{R}-Y_i)_{\gamma(t)}, e_i(0)\right\rangle \\
&= \frac{1}{t}\sum_{i=1}^{n}\left\langle A_{te_n(0)}^{-1}\nabla_{Y_i}\mathcal{R}-e_i(0), e_i(0)\right\rangle \\
&= \frac{1}{t}\sum_{i,j,k,l=1}^{n}\bar{a}_{jk}(t)a_{il}(t)\left\langle\nabla_{e_l(t)}\mathcal{R}, e_j(t)\right\rangle\langle e_k(0), e_i(0)\rangle - \frac{n}{t} \\
&= \frac{1}{t}\sum_{i=1}^{n}\left\langle\nabla_{e_i(t)}\mathcal{R}, e_i(t)\right\rangle - \frac{n}{t} \\
&= \frac{1}{t}(\operatorname{div}\mathcal{R})_{\gamma(t)} - \frac{n}{t}.
\end{aligned}
$$

But

$$(\mathcal{R}\log\mathcal{J})_{\gamma(t)} = t\frac{d}{dt}\log\mathcal{J}(t),$$

so $\operatorname{div}\mathcal{R} = \mathcal{R}\log\mathcal{J} + n$. This is what was left for proving the proposition.

EXERCISE 8.3.6. Using Dynkin's formula, as in the Euclidian case, show that

$$E[\zeta_r] = \frac{r^2}{n} - \frac{1}{n}E\left[\int_0^{\zeta_r(p)}(\mathcal{R}\log\mathcal{J})(p(s))ds\right].$$

The formula of the previous exercise will be the basis for estimating the expected value of ζ_r.

2. *Approximating* $\mathcal{R}\log\mathcal{J}$. Before moving forward, we introduce a few definitions.

EXERCISE 8.3.7. Let $(M, \langle\cdot,\cdot\rangle)$ be a Riemannian manifold. Let $e_1, \ldots, e_n$ be an orthonormal basis of T_qM, for some $q \in M$, and define for $u, v \in T_qM$:

$$\operatorname{Ric}_q(u, v) := \sum_{i=1}^{n}\langle R(e_i, u)v, e_i\rangle.$$

Show that the definition does not depend on the choice of orthonormal basis and that $q \mapsto \operatorname{Ric}_q$ is a smooth tensor field. Also define

$$S(q) := \sum_{i=1}^{n}\operatorname{Ric}_q(e_i, e_i).$$

Show that $S(q)$ does not depend on the choice of orthonormal basis and $q \mapsto S(q)$ is a smooth function on M.

For a given unit vector $u \in T_qM$, $\mathrm{Ric}_q(u) := \mathrm{Ric}_q(u,u)$ is called the *Ricci curvature* in the direction u, while $S(q)$ is the *scalar curvature* at q.

In this section we will show that the following formula holds for every unit vector $v \in T_qM$:

$$(\mathcal{R}\log\mathcal{J})(\exp_q tv) = -\frac{t^2}{3}\mathrm{Ric}_q(v) + O(t^3). \tag{8.1}$$

It will be convenient to introduce a vector field $Z_i(t)$ along the geodesic $\exp_q(tv)$ defined as follows: Let $e_1,\dots,e_n$ be an orthonormal basis at q such that $e_n = v$. Then

$$Z_i(t) := \left.\frac{\partial}{\partial s}\right|_{s=0} \exp_q(t(v+se_i)).$$

EXERCISE 8.3.8. Show that $Z_i(t) = t(d\exp_q)_{tv}e_i$. Check that $Z_i(0) = 0$ and $\frac{\nabla Z_i}{dt}(0) = e_i$. If $e_i(t)$ denotes, for each i, the parallel translation of e_i along $\exp_q tv$, show that

$$\det(\langle Z_i(t), e_j(t)\rangle) = t^n\mathcal{J}(\exp_q tv).$$

EXERCISE 8.3.9. Show that $Z_i(t)$ satisfies the following differential equation:

$$\frac{\nabla^2 Z_i}{dt^2}(t) = R(e_n(t), Z_i(t))e_n(t).$$

We give here some indications for solving the exercise. Define

$$\Phi(s,t) = \exp_q(tv + tse_i).$$

Introduce the vector fields (on the surface parametrized by Φ):

$$V(s,t) := \frac{\partial\Phi}{\partial s}(s,t),\ W(s,t) := \frac{\partial\Phi}{\partial t}(s,t).$$

Check that V and W commute and that $V(0,t) = Z_i(t), W(0,t) = e_n(t)$. Since $t \mapsto \Phi(s,t)$ is a geodesic for each s, then $\frac{\nabla W}{dt} = 0$. We now write:

$$\begin{aligned}\frac{\nabla^2 Z_i}{dt^2}(t) &= \left(\frac{\nabla}{dt}\frac{\nabla}{dt}\frac{d}{ds}\Phi(s,t)\right)_{s=0}\\ &= \left(\frac{\nabla}{dt}\frac{\nabla}{ds}\frac{d}{dt}\Phi(s,t)\right)_{s=0}\\ &= \left(\frac{\nabla}{dt}\frac{\nabla}{ds}W(s,t) - \frac{\nabla}{ds}\frac{\nabla}{dt}W(s,t)\right)_{s=0}\\ &= R(e_n(t), Z_i(t))e_n(t).\end{aligned}$$

The above equation characterizes a *Jacobi vector field*.

To obtain the expansion 8.1 we begin by finding the Taylor series of

$$\beta_{ij}(t) := \langle Z_i(t), e_j(t)\rangle.$$

The key point to observe is that, since $e_n(t)$ is parallel along $\exp_q tv$:

$$\beta_{ij}^{(k)}(0) = \left\langle \frac{\nabla^k Z_i}{dt^k}(0), e_j(0)\right\rangle.$$

Now the Jacobi equation provides a means to obtain the succecive derivatives. We write below the first few. The following notation will be used:

$$\begin{aligned}\frac{\nabla R}{dt}(X,Y)Z &:= (\nabla_{e_n} R)(X,Y)Z \\ &:= \frac{d}{dt}R(X,Y)Z - R\left(\frac{\nabla X}{dt}, Y\right)Z - R\left(X, \frac{\nabla Y}{dt}\right)Z - R(X,Y)\frac{\nabla Z}{dt}.\end{aligned}$$

$(\nabla_{X_1} R)(X_2, X_3)X_4$ actually defines a tensor field of type $(4,1)$. The derivatives, up to order 4, are now as follows:

$$\begin{aligned}\beta_{ij}(0) &= 0; \\ \beta'_{ij}(0) &= \delta_{ij}; \\ \beta''_{ij}(0) &= 0; \\ \beta'''_{ij}(0) &= \langle R(v, e_i(0))v, e_j(0)\rangle; \\ \beta^{(4)}_{ij}(0) &= 2\langle (\nabla_v R)(v, e_i(0))v, e_j(0)\rangle.\end{aligned}$$

Therefore, we have:

$$\beta_{ij}(t) = t\delta_{ij} + \frac{t^3}{6}\langle R(v, e_i(0))v, e_j(0)\rangle + O(t^4).$$

EXERCISE 8.3.10. Show the following sequence of identities:

$$\begin{aligned}(\mathcal{R}\log\mathcal{J})(\exp_q tv) &= -n + t\frac{d}{dt}\log\det\beta \\ &= -n + \operatorname{Trace}\left(\beta^{-1}t\frac{d}{dt}\beta\right) \\ &= -\frac{t^2}{3}\operatorname{Ric}_q(v) + t^3 W_q(v) + O(t^4).\end{aligned}$$

In the last line, W_q is a cubic form in v, whose exact expression won't be needed.

EXERCISE 8.3.11. A Riemannian manifold $(M, \langle\cdot,\cdot\rangle)$ is said to be a (locally) *symmetric space* if the following property holds: For each $q \in M$ and each nonzero $v \in T_qM$, there exists a (locally defined, around q) isometry $h : M \to M$ such that $dh_q v = -v$. Show that the curvature tensor of a locally symmetric space has the property $\nabla R = 0$.

Examples of symmetric spaces are spheres, hyperbolic spaces (real, complex, quaternionic), and projective spaces (real, complex, quaternionic). Notice that for a locally symmetric space the same expression shown above for the Taylor expansion of $\beta_{ij}(t)$ (up to third order) has an error of the order $O(t^5)$.

Here is where we stand: Writting $u(s) := (\exp_q)^{-1}(p(s))$, then

$$E[\zeta_r] = \frac{r^2}{n} + \frac{1}{3n}E\left[\int_0^{\zeta_r(p)} \left\{\operatorname{Ric}_q(u(s)) + W_q(u(s)) + O(|u(s)|^4)\right\} ds\right].$$

Therefore, to move towards our goal, we need to obtain the estimate:

$$E\left[\int_0^{\zeta_r(p)} \{\operatorname{Ric}_q(u(s)) + W_q(u(s))\}\, ds\right] = \frac{r^4}{2n(n+2)}S(q) + o(r^4).$$

In order to better explain how this estimation will proceed, we introduce the following two processes on T_qM. Fix an orthonormal frame $\sigma : \mathbb{R}^n \to T_qM$ Then the first process, $\beta : \mathcal{P}_0(\mathbb{R}^n) \to \mathcal{P}_0(T_qM)$ is simply ordinary Brownian motion on T_qM, obtained by

$$\omega \in \mathcal{P}_0(\mathbb{R}^n) \mapsto \sigma \circ \omega \in \mathcal{P}_0(T_qM).$$

The second process is essentially Brownian motion on M, brought to T_qM via the exponential map. More precisely, we introduce $\beta^M : \mathcal{P}_0(\mathbb{R}^n) \to \mathcal{P}_0(T_qM)$ such that

$$\beta^M(t, \omega) = (\exp_q)^{-1}(p(t, \omega)).$$

Notice that $\beta^M(t)$ is defined (at least) up to stopping time $\zeta_r(p)$, where r is small enough for $\exp_q$ to be a diffeomorphism from $B(0, r) \subset T_qM$ to the ball of radius r, center q, in M.

The following lemma will be needed. It involves the class $\mathcal{C}_{m,r_0,C}(\mathbb{R}^n)$, for $r_0 > 0, m \geq 1$, consisting of continuous functions $\psi : \mathbb{R}^n \to \mathbb{R}$ such that $\psi(0) = 0$ and

$$\sup_{x,x' \in B(0,r), x \neq x'} \frac{|\psi(x) - \psi(x')|}{|x - x'|} \leq C r^{m-1}$$

for all $r \in (0, r_0]$ and some $C < \infty$. In the lemma, $\zeta_r(\gamma)$ is defined as before, except that we are now considering the first exit time from the ball $B(0, r)$ in T_qM.

LEMMA 8.3.12. *There exists an $r_0 > 0$ and $K < \infty$ such that, for all $\psi \in \mathcal{C}_{m,r_0,C}(\mathbb{R}^n)$, we have*

$$\left| E\left[\int_0^{\zeta_r(\beta)} \psi(\beta(s))ds\right] - E\left[\int_0^{\zeta_r(\beta^M)} \psi(\beta^M(s))ds\right]\right| \leq KCr^{m+4} \log \frac{1}{r}$$

whenever $0 < r < r_0$.

The point of the lemma is that it will permit us to use ordinary Brownian motion on T_qM to estimate the expected value of ζ_r for the Riemannian Brownian motion, for sufficiently small r. We will discuss its proof later in the section. The proof will be given later.

Our next step is to find

$$E\left[\int_0^{\zeta_r(\beta)} (\mathrm{Ric}_q(\beta(s)) + W_s(\beta(s))ds\right].$$

Since the Wiener measure is invariant under the transformation

$$\gamma \mapsto -\gamma, \gamma \in \mathcal{P}_0(\mathbb{R}^n),$$

the cubic term can be disregarded.

EXERCISE 8.3.13. Check that Ric_q is a symmetric quadratic form on T_qM. Therefore, it is diagonalizable. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues associated to an orthonormal basis $e_1, \ldots, e_n$ of eigenvectors of Ric_q. Show that

$$\mathrm{Ric}_q(v) = \sum_{i=1}^{n} \lambda_i \langle v, e_i \rangle^2$$

and that $S(q) = \lambda_1 + \cdots + \lambda_n$.

Given the previous exercise, all that is left is show that

$$E\left[\int_0^{\zeta_r(\beta)} \langle \beta(s), e_i \rangle^2 ds\right] = \frac{r^4}{2n(n+2)}.$$

Equivalently, since $|w|^2 = \langle w, e_1\rangle^2 + \cdots + \langle w, e_n\rangle^2$, we want to prove:

$$E\left[\int_0^{\zeta_r(\beta)} |\beta(s)|^2 ds\right] = \frac{r^4}{2(n+2)}.$$

Let $f(x) = |x|^4$. Then $(\Delta f)(x) = 4(2+n)|x|^2$. Using Dynkin's formula we get

$$\begin{aligned} r^4 &= E[|\beta_{\zeta_r(\beta)}|^4] \\ &= 2(2+n)E\left[\int_0^{\zeta_r(\beta)} |\beta(s)|^2 ds\right]. \end{aligned}$$

3. *Proof of Lemma 8.3.12.*

Bibliography

[1] *. *,